Spring 2002

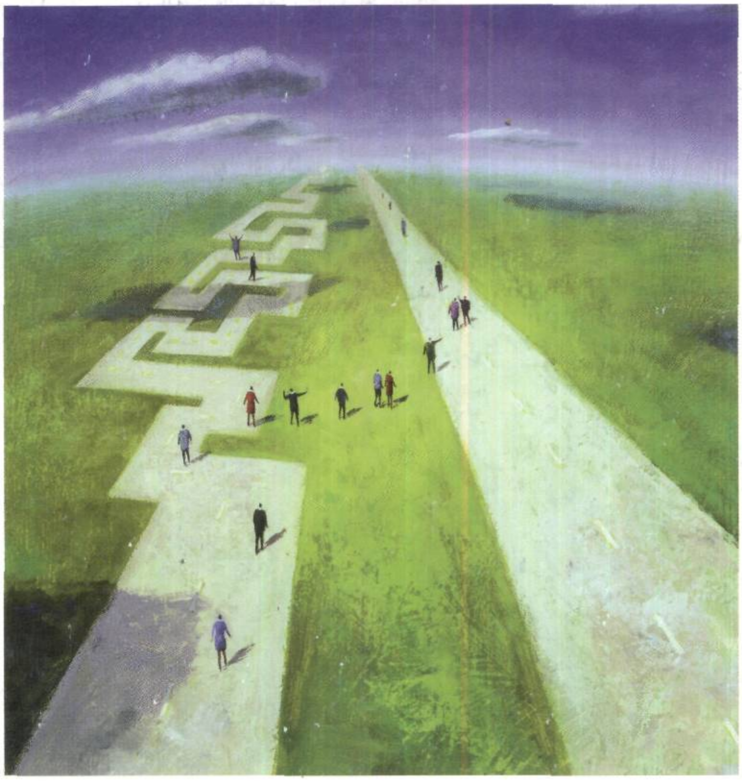
EDUCATION next

A JOURNAL OF OPINION AND RESEARCH

New Paths to Teaching

Should education schools control certification?

MARY DIEZ • JAMES FRASER • FREDERICK HESS



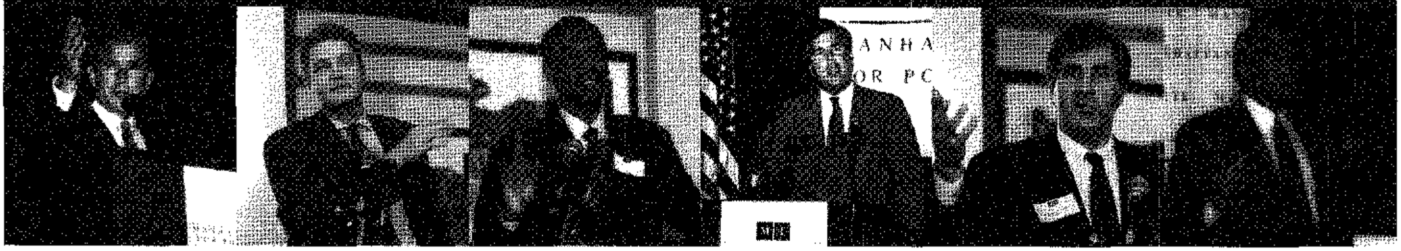
IN THIS ISSUE:

Phi Delta Kappa Cooks the Questions, The Mystery of Good Teaching, Teach for America, Measuring School Progress, Debating the New Federal Role

CONTRIBUTORS: TERRY MOE, DAN GOLDHABER, DAVID RUENZEL, KATE WALSH, THOMAS KANE, MARGARET RAYMOND, RICHARD ELMORE, ANDREW ROTHERHAM, JEFFREY JONES

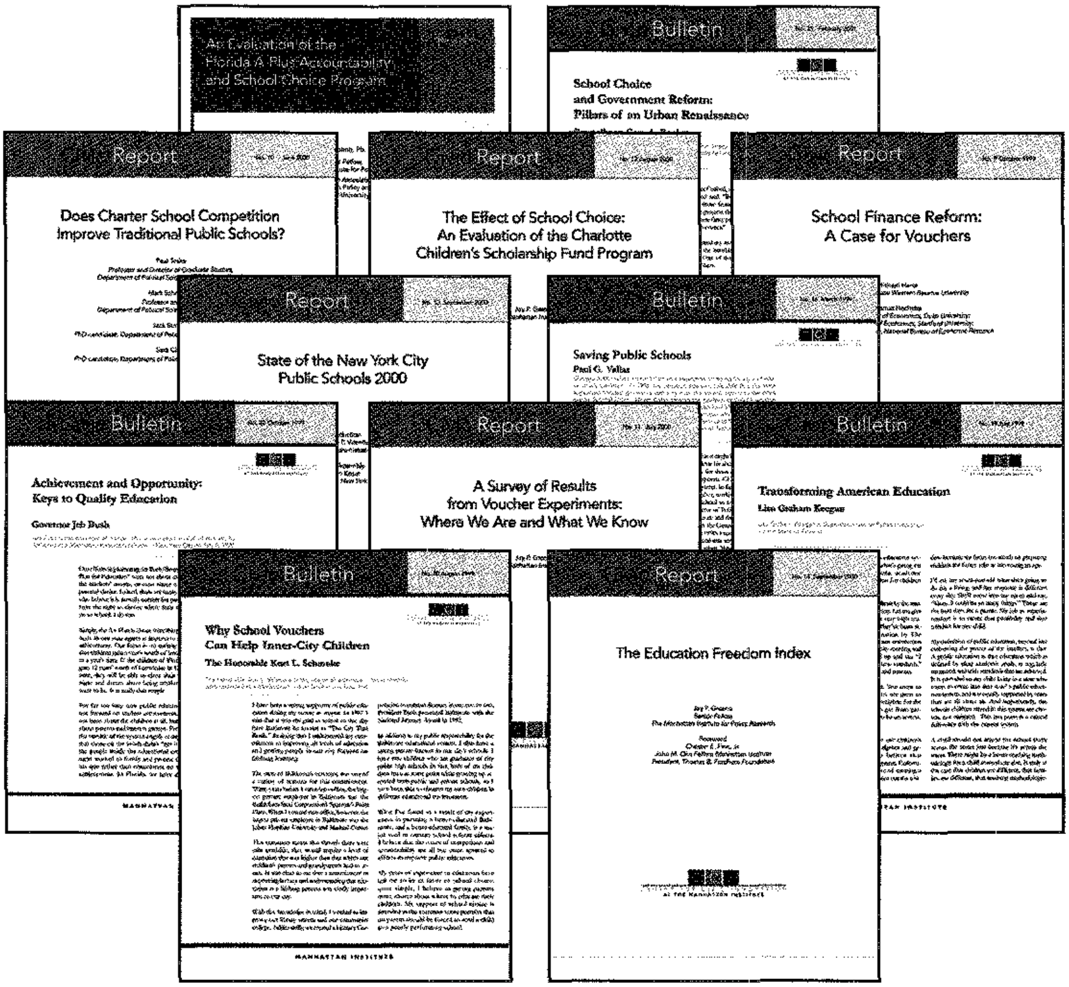
www.educationnext.org / $7.00

Making the case for education reform . . .



U.S. President George W. Bush **Mayor John Norquist** **Reverend Floyd Flake** **Governor Jeb Bush** **Mayor Bret Schundler** **U.S. Secretary of Education Rod Paige**

. . . one brick at a time.



An Evaluation of the Florida A-Plus Accountability and School Choice Program

Bulletin
School Choice and Government Reform: Pillars of an Urban Renaissance

Report
Does Charter School Competition Improve Traditional Public Schools?

Report
The Effect of School Choice: An Evaluation of the Charlotte Children's Scholarship Fund Program

Report
School Finance Reform: A Case for Vouchers

Report
State of the New York City Public Schools 2000

Bulletin
Saving Public Schools
Paul G. Vallas

Bulletin
Achievement and Opportunity: Keys to Quality Education
Governor Jeb Bush

Report
A Survey of Results from Voucher Experiments: Where We Are and What We Know

Bulletin
Transforming American Education
Lisa Graham Keegan

Bulletin
Why School Vouchers Can Help Inner-City Children
The Honorable Kurt L. Schmoke

Report
The Education Freedom Index

MANHATTAN INSTITUTE FOR POLICY RESEARCH

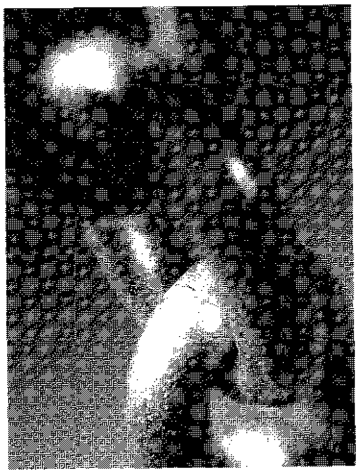
"turning intellect into influence"

www.manhattan-institute.org

52 Vanderbilt Avenue • New York, NY 10017 • phone 212.599.7000 • fax 212.599.3494

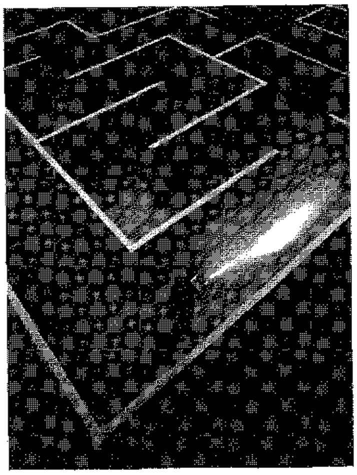
ELECTRONIC REPRODUCTION PROHIBITED

VOLUME 2, NUMBER 1
EDUCATION next
A JOURNAL OF OPINION AND RESEARCH



7

29



42

forum

7 Teaching's Gatekeepers
Should education schools lose their monopoly?
The Certification Connection by MARY E. DIEZ *A Tenuous Hold* by JAMES W. FRASER
Break the Link by FREDERICK M. HESS

29 The Feds Step In
Will the new accountability regime spur improvement?
Unwarranted Intrusion by RICHARD F. ELMORE *A New Partnership* by ANDREW ROTHERHAM

features

42 Tortuous Routes
Wanted: a real alternative to traditional teacher certification by DAVID RUENZEL

50 The Mystery of Good Teaching
Surveying the evidence on student achievement and teachers' characteristics by DAN GOLDHABER

research

56 Randomly Accountable
Test scores and volatility by THOMAS J. KANE, DOUGLAS O. STAIGER, & JEFFREY GEPPERT

research

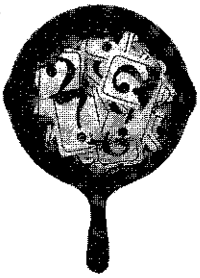
62 **Teach for America**
The first evidence on classroom performance
by MARGARET RAYMOND & STEPHEN FLETCHER

62

check the facts

70 **Cooking the Questions?**
The 33rd Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools
by Lowell C. Rose & Alec M. Gallup
CHECKED by TERRY M. MOE

79 **Positive Spin**
The evidence for teacher certification.
CHECKED by KATE WALSH



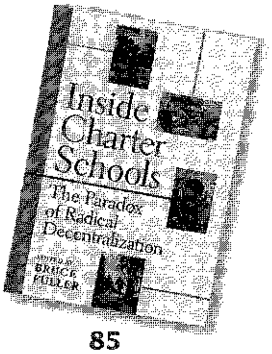
70

from the editors

3 **Certifying the Intangible**

correspondence

4 *A-plus for Florida?*



85

book review

85 **Inside Charter Schools: The Paradox of Radical Decentralization**
Edited by Bruce Fuller
REVIEWED by PATRICK J. WOLF

education matters to me

88 **Low Expectations**
by JEFFREY JONES

88

COVER ILLUSTRATION *by Andrew Judd/Masterfile*

VOLUME 2, NUMBER 1

EDUCATION next

A JOURNAL OF OPINION AND RESEARCH

Editor-in-Chief	Paul E. Peterson
Senior Editor	Chester E. Finn Jr.
Executive Editors	Frederick M. Hess Marci Kanstoroom
Articles Editor	Tyce Palmaffy
Managing Editor	Carol Peterson
Research Editor	Martin West
Design Director	John Sizing
Consulting Editors	Patricia Baker, Jay P. Greene
Web Editor	Bridgitt Ahern
Circulation	Jennifer Presley
Editorial Board	John E. Chubb, Williamson Evers, Eric A. Hanushek, Paul Hill, E. D. Hirsch Jr., Caroline M. Hoxby, Terry M. Moe, Diane Ravitch, Herbert Walberg

---

Publisher	Hoover Institution, Leland Stanford Junior University

---

Additional Sponsoring Institutions	Program on Education Policy and Governance, Harvard University Thomas B. Fordham Foundation Manhattan Institute for Policy Research

---

Subscriptions One-year subscription (four issues), $20 in the United States; $30 in Canada; $40 elsewhere. To subscribe or to submit change of address information, call us at (650) 723-3373; fax us at (650) 723-8626; visit us at www.educationnext.org; or email us at hooverpress@hoover.stanford.edu. You may also write to *Education Next* Subscriptions, Hoover Institution, Stanford University, Stanford, CA 94305-6010. Cover price is $7/issue. ISSN 1532-5148.

Advertising For information on advertising, contact Thomas Polseno, Program on Education Policy and Governance, Kennedy School of Government, Harvard University, Cambridge, MA 02138, (617) 495-6954.

Letters We welcome letters about our contents. Write to Letters to the Editor, *Education Next*, 226 Littauer North Yard, 1875 Cambridge Street, Cambridge, MA 02138; fax to (617) 496-1507; or email Editor_EM@latte.harvard.edu.

Editorial Queries We consider unsolicited contributions. Send manuscript submissions as email attachments to Editor_EM@latte.harvard.edu. Pieces sent by regular mail go to Carol Peterson, *Education Next*, 226 Littauer North Yard, 1875 Cambridge Street, Cambridge, MA 02138.

Opinions expressed in *Education Next* are those of the authors and do not necessarily reflect those of the Editorial Board of *Education Next*, the Board of Trustees of Leland Stanford Junior University, or the sponsors of *Education Next*.

Visit us at http://www.educationnext.org, where unabridged essays can be found.

© by the Board of Trustees of Leland Stanford Junior University, 2002

Certifying the Intangible

No one really knows what makes a good teacher

Do we need good teachers? Don't be silly. Of course we do. We can all recall a teacher who made a big difference in our lives. And now we have research, as reviewed in Dan Goldhaber's **Feature** essay, which shows more clearly than ever before that the quality of the teacher has more influence over student learning than anything else at school.

But what makes for a good teacher? Aye, there's the rub. Certainly, it's nothing easily put into a box and counted. Whether a teacher has a master's or other advanced degree appears to count only in science and math. A few years of experience in the classroom makes some difference, but less than one would think. A teacher's command of the English language seems to be an important predictor of a teacher's effectiveness, at least in the United States. But even this, together with all the other measurable elements, leaves more than 90 percent of a teacher's influence unexplained.

> Teacher quality has more influence over student learning than anything else at school.

When so little is known with certainty, how should we recruit teachers? Most states presently insist that teachers take a specific curriculum taught in schools of education in order to be certified to work in public school classrooms. Unless one follows this well-worn path or gets an emergency exemption, David Ruenzel tells us in his **Feature** essay, the maze that a prospective teacher must negotiate is filled with dead ends. Yet the hoops and hurdles of certification persist—despite the fact that alternative paths have considerable promise. So show Margaret Raymond and Stephen Fletcher in their careful **Research** on the classroom performance of Teach for America recruits in Houston. Even more to the point, Kate Walsh **Checks the Facts** underlying the claims that certification boosts teacher performance, and demonstrates that traditional education coursework, typically required by certification standards, has never been shown—even by its most ardent champions—to have a payoff in the classroom.

Yet certification laws still have their defenders, including Mary Diez, who contends in our **Forum** on teacher certification that the new standards for teacher education will ensure that teachers are "well equipped to work effectively with learners." In response, James Fraser and Frederick Hess propose alternatives that would allow superintendents and principals, rather than state boards and schools of education, to decide who is qualified to teach.

Key to granting such freedoms to local educators is finding legitimate ways to hold them accountable for their results. The new federal law demands that schools meet certain performance goals or risk sanctions as severe as having their entire staffs fired and forced to reapply for their jobs, a measure known as "reconstitution." In enacting it, Congress took note of the **Research** presented here by Thomas Kane and his colleagues, which shows just how complicated it is to judge schools by changes in their test scores. That problem also gives bite, in our **Forum** on accountability, to Richard Elmore's tough critique of the bipartisan federal initiative. In response, Andrew Rotherham tells us not to rush to judgment too quickly. Testing, he argues, has had positive consequences in some states, and the wrinkles in the system can be ironed out in the years ahead.

Phi Delta Kappa (PDK), however, isn't even trying to iron out the wrinkles in its widely cited surveys of America's attitudes toward the public schools. Terry Moe's powerful **Check the Facts** shows that PDK, using the venerable Gallup label, routinely fixes its survey questions to get the answers it wants, at least on the issue of school vouchers. PDK neglected to change its format, even after Gallup itself proved that its survey questions on vouchers were biased. We imagine that Mark Twain would use PDK's results as Exhibit A if his famous line on lies and statistics were ever on trial.

-THE EDITORS

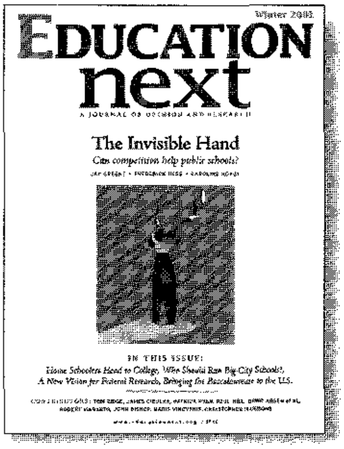
MISSION STATEMENT In the stormy seas of school reform, this journal will steer a steady course, presenting the facts as best they can be determined, giving voice (without fear or favor) to worthy research, sound ideas, and responsible arguments. Bold change is needed in American K-12 education, but *Education Next* partakes of no program, campaign, or ideology. It goes where the evidence points.

A-Plus for vouchers?

In "The Looming Shadow" (*Research*, Winter 2001), Jay P. Greene of the Manhattan Institute examines whether the threat of vouchers under Florida's A-Plus program forced the state's failing schools to improve. The A-Plus program is essentially a top-down accountability system with a voucher add-on. The state grades schools from A to F based on their test-score performance. If schools labeled F fail to improve, their students become eligible for vouchers.

Greene found that the test scores of failing schools did increase. This leads him to jump to the conclusion that the threat of vouchers was the primary reason. However, how Greene can distinguish, based on Florida data alone, the effect of the voucher threat from the effect of being designated a failing school escapes me.

Greene makes an attempt at distinguishing between these two effects by narrowing the comparison to the highest-performing F schools and the lowest-performing D schools (so that the only real difference between the schools was whether they faced the threat of vouchers or not). He notes that the F schools were provided with $600 more per pupil than the D schools, but he finds that taking those resources into account would not change the conclusion. The basic problem remains, however. Given the increased public scrutiny and the stigma associated with being labeled a failing school, we should expect such schools to work hard at improving regardless of whether their students are given the option of a voucher.



Winter 2001

EDUCATION next

The Invisible Hand

Can competition help public schools?

IN THIS ISSUE:

Home Schoolers Head to College, Who Should Run Big-City Schools?, A New Vision for Federal Research, Bringing the Baccalaureate to the U.S.

North Carolina has a similar accountability system, in that low-performing schools are publicly identified as failing. My Duke University colleague Beth Glennie and I have found that schools that were designated as failing increased their performance the following year more than all the other categories of schools. Moreover, using Greene's methodology of narrowing the comparison to the highest performing schools in the failing group and the lowest-performing in the next group, we still find that failing schools improved more. In other words, schools in North Carolina exhibited the same pattern of improvement, yet they faced no threat of vouchers. Instead their performance reflected the response of failing schools to some combination of increased scrutiny, the shame of being labeled a failure, and intervention from state assistance teams.

Given that similar factors are at work in Florida's accountability system, I suspect that most, if not all, of the improvements in school performance in that state's failing schools are attributable to the state's administered accountability system, not to the voucher component of that program.

HELEN F. LADD
Duke University
Durham, North Carolina

Jay P. Greene Responds: Helen Ladd is correct in saying that the Florida A-Plus program is an "accountability system with a voucher add-on," or, as I put it more forcefully in my article, "an accountability system with teeth." She questions whether vouchers were the real "teeth" of the A-Plus program, contending that public shame or the threat of reconstitution may have been just as effective, if not more so, as the prospect of vouchers.

Ladd bolsters her claim with evidence from North Carolina that schools labeled as failing exhibited similar patterns of improvement, but without the threat of vouchers. However, her study and others like it neglect to establish that the gains realized by failing schools facing alternative sanctions represent real improvement in student learning. By contrast, my study validated the state test by correlating the state test results with the results of low-stakes national test results. Without establishing the validity of state testing results, it is impossible to know whether the gains made by failing schools in other states were as large as those realized by failing schools that faced the prospect of vouchers in Florida.

Even if sanctions other than vouchers inspired school improvement, my study's findings would not be undermined in the least. The point of my study was not that vouchers are the only effective sanction. My point was that vouchers were an effective sanction in Florida and that schools needed to have incentives in addition to resources in order to improve. Helen Ladd seems to agree that sanctions are effective at inspiring school improvement, suggesting that we agree on my fundamental point about the crucial role of incentives in school reform.

Old choices

A common shortcoming in research and commentary on school choice is the failure to recognize the extent to which school choice already exists. As a result, a typical theme is that there is "little evidence" on school choice, so that only "preliminary" and "tentative" conclusions can be reached.

This assumption underlies each of the articles in your "When Schools Compete" Forum (Winter 2001). The contributors discuss two limited forms of choice in K–12 education—vouchers and charter schools—when in fact a large share of the population has always exercised one or another form of choice. As Richard Elmore and Bruce Fuller explain in *Who Chooses? Who Loses?* "Choice is everywhere in American education. It is manifest in the residential choices made by families . . . [and] when families, sometimes at great financial sacrifice, decide to send their children to private schools. . . . In all instances, these choices . . . are strongly shaped by the wealth, ethnicity, and social status of parents and their neighborhoods."

Furthermore, for decades, higher education in America has flourished as a wide-open system of tax-supported educational choice involving all manner of public and private institutions. Indeed, many failing K–12 public school systems exist side-by-side with networks of thriving public and private universities. Is this not evidence of competition's long-term success in the education sphere?

Given the widespread existence of choice and competition in K–12 education, it is demonstrably wrong to suppose, as Forum contributor Frederick M. Hess does ("The Work Ahead"), that "Efforts to cultivate competition may thus foster a culture of schooling that is alien to our educational heritage and may create an incentive structure that distorts educational priorities."

Likewise, in "Finishing Touches," Robert Maranto states, "The animating theory of school choice has always been that it will not only serve as an escape hatch from dysfunctional public schools but also will spark public schools to improve. Thus far this theory remains mostly untested." For many low-income parents, choice always has been about equity, about having what other parents take for granted. The idea that the effect of choice is "mostly untested" again highlights the core problem with these three articles.

HOWARD FULLER
GEORGE A. MITCHELL
Marquette University
Milwaukee, Wisconsin

Ripple effect

As the articles by Caroline Minter Hoxby ("Rising Tide," *Research*, Winter 2001), Robert Maranto, and Jay P. Greene show, incentives have the power to change behavior in ways that improve both private and public schools.

When Milwaukee was only a few years into its choice program, the school board did something it had never done before—it enacted standards for children to attain in math and reading.

Likewise, a group of parents in Minnesota tell a story of how they long wanted a Montessori elementary school in their community. They tried to cajole, petition, and urge local officials into creating such a program, without success. Then came Minnesota's charter school law, the nation's first, in 1991. Lo and behold, as these parents were putting the finishing touches on their proposal, the school district offered them the opportunity to create a new Montessori elementary school.

Ask the superintendents and school boards about the changes they enacted, and they'll tell you they were planning to do it anyway. That's what the principals and school leaders said in Florida, when they reacted to the state's voucher program with an unprecedented campaign to turn around long-failing schools and introduced new reading programs.

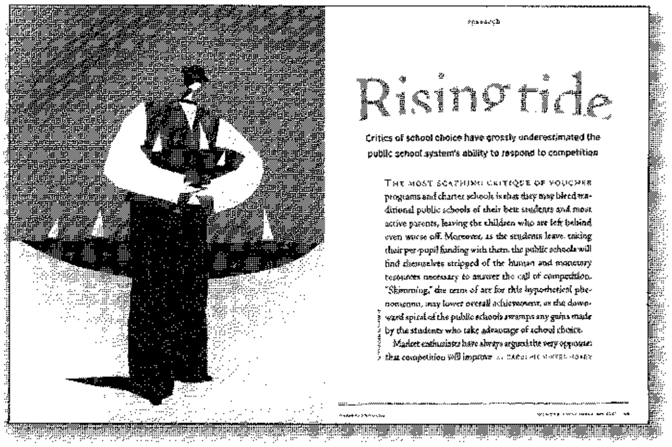
JEANNE ALLEN
Center for Education Reform
Washington, D.C.

Cloning Houston

Praise for the Houston school district comes at the expense of cities that have taken similar steps forward. Paul T. Hill ("Digging Deeper," *Feature*, Fall 2001) gives a passing nod to Chicago and Community District 2 in New York, but the progress in Houston is becoming less rare than Hill suggests. New data from Charlotte and Sacramento indicate comparable advances. A second tier of cities—including Fort Worth, Long Beach, Norfolk, Boston, Indianapolis, Louisville, San Diego, and Minneapolis—are posting gains that may be short of Houston's, but are exceeding their respective state averages. Moreover, cities like Baltimore and Cleveland are transforming themselves and have shown impressive spikes on state tests in recent years.

Hill should also know that urban schools are not governed by gypsy superintendents who wander the nation in search of leadership positions. In fact, only nine of the superintendents in the nation's largest urban school systems have ever been superintendents in another major city (which is less than the number of nontraditional superintendents now running major school systems, of which there are 11). Only one—Boston superintendent Thomas Payzant—has headed two other major systems, and he can hardly be described as nomadic.

MICHAEL CASSERLY
Council of the Great City Schools
Washington, D.C.



research

Rising tide

Critics of school choice have grossly underestimated the public school system's ability to respond to competition

THE MOST SCATHING CRITIQUE OF VOUCHER programs and charter schools is that they may bleed traditional public schools of their best students and most active parents, leaving the children who are left behind even worse off. Moreover, as the students leave, taking their per-pupil funding with them, the public schools will find themselves stripped of the human and monetary resources necessary to answer the call of competition. "Skimming," the term of art for this hypothetical phenomenon, may lower overall achievement, as the downward spiral of the public schools swamps any gains made by the students who take advantage of school choice.

Market enthusiasts have always argued the very opposite: that competition will improve

Civic education

David Campbell argues that one of the primary goals of a public education is to teach children "the ability to deliberate" in a setting of "mutual respect among persons." But fostering mutual respect for persons of different ethnic, racial, and cultural backgrounds is not part of our secular concept of American democracy. Moral tolerance is a normative, private virtue, not a public good. Perhaps this explains why students at religious schools score higher on measures of civic participation (volunteering in the community) than public school children. The more fruitful avenue of research, based on what Campbell has told us, is to look at how private secular schools are serving our political system.

RACHEL M. MCCLEARY
Harvard University
Cambridge, Massachusetts

David Campbell's article "Bowling Together" (*Research*, Fall 2001) struck a chord with me as both an educator and a father. For 15 years as a college president, I watched each incoming freshman class exhibit less civic engagement and more cynicism toward political processes. Campbell's conclusion, that religious schools, and Catholic schools especially, have a favorable effect on developing civic participation, came as no surprise to me. With my own children in both public and religious schools, I have seen significantly more emphasis on understanding and being active in government in the religious schools, including yearlong studies in preparation for major trips to Washington, D.C., and the state capital.

It is a shame that the pendulum has swung so far toward the separation of church and state that we have difficulty acknowledging the role church-related schools might play in strengthening education in America. As president of Pepperdine University, when I was asked whether I thought there was something "artificial" about having a religious emphasis in an academic environment, I generally responded that it seemed more artificial to me to look at subjects from every conceivable point of view except the spiritual. The big difference in most religious schools is not that they indoctrinate, but that they raise topics such as moral, ethical, civic, and spiritual matters that teachers in public schools cannot or dare not address.

The big difference in most religious schools is not that they indoctrinate, but that they raise topics such as moral, ethical, civic, and spiritual matters that teachers in public schools cannot or dare not address.

Nevertheless, I'm not certain that Campbell's use of community service as a measure of civic participation is a good predictor for voting patterns and other forms of adult civic engagement. The current pattern on college campuses, where students can both volunteer and vote, is that they do the former and not the latter. Community service is at all-time highs on college campuses, voting at all-time lows. As the old bumper sticker says, we need citizens who will think globally as well as act locally.

DAVID DAVENPORT
Hoover Institution
Palo Alto, California

Promoting patriotism

Diane Ravitch's lament that public schools no longer teach a common culture (see "Ex Uno Plures," *Forum*, Fall 2001) is even more relevant after the events of September 11.

Our greatest challenge is to prevent more terrorism, but our greatest opportunity is to harness the renewed sense of civic pride to make our country permanently stronger. This will not be easy. Robert Putnam, author of *Bowling Alone*, says that national tragedies have often been followed by an upward spike in civic engagement. Only in the case of Pearl Harbor did it last. Americans who lived through December 7, 1941, changed their way of living; they grew victory gardens, bought war bonds, rationed gas, and even welcomed hitchhiking servicemen, getting them where they needed to go. World War II, building as it did on the suffering of the Great Depression, created a new civic mindedness.

One way that public schools can make the changes of September 11 more lasting is to start each school day with the Pledge of Allegiance, followed by a student or faculty member explaining "what it means to be an American" for three minutes. A student of Kurdish descent might talk about why Americans don't persecute all Muslims just because some Muslims are terrorists. A teacher might explain why we say "In God We Trust," but we don't trust government with God. A principal might discuss why some civil liberties extend even to people who try to kill us. Ravitch reminds us that the historical mission of the common school included helping children learn what it means to be an American. Teaching American values in the context of reciting the pledge would honor the victims and heroes of September 11 as well as honor the reason why public schools exist in the first place.

LAMAR ALEXANDER
Harvard University
Cambridge, Massachusetts

TEACHING's gatekeepers

Should education schools continue to be the only route to teaching in the public system?

For the past half-century, states have guaranteed university- and college-based schools of education a steady stream of students and revenues by certifying their graduates, and only their graduates, to teach in public schools. However, as the teacher shortage worsens, especially in urban areas, and as evidence on the success of alternative certification programs like Teach for America grows, policymakers are wondering whether the current licensure system needlessly restricts the pool of potential teachers. As a result, alternative programs that speed candidates into the classroom, allowing them to teach full time while working toward a credential, are proliferating in cities like New York and Washington, D.C. Meanwhile, a separate reform movement, spearheaded by the National Commission on Teaching & America's Future, is trying to bolster education schools while insisting that states require all teachers to pass through them. Which approach will attract more talented, skilled candidates to the teaching profession?

PAGE 8
Mary E. Diez *praises the new standards for teacher training*

PAGE 16
James W. Fraser *wants education schools to relinquish their monopoly—for their own good*

PAGE 22
Frederick M. Hess *designs a competitive certification system*

THE certification connection

Licensure ought to guarantee that every classroom comes equipped with a skilled, knowledgeable teacher. The new performance standards for teachers are making that possible

I recently came across a flyer from the National Private Schools Association offering, among other things, certification for private school teachers. Intrigued, I went to the association's website and discovered these requirements: provide information, all self-reported, on your academic background, teaching experience, and character; obtain a reference from an employer or colleague who will nominate you for certification (although no *criteria* for such nomination are provided); and pay $50 (your school also must pay an annual fee of $75 for you to be eligible).

While the existing system of licensure for public school teachers is more rigorous than simply rubber-stamping someone's self-report on competence, most thoughtful people agree that the system is in need of a major overhaul. As a result, the processes and institutions that license teachers are changing, and many of the changes promise to ensure that teachers enter the classroom well equipped to work effectively with learners.

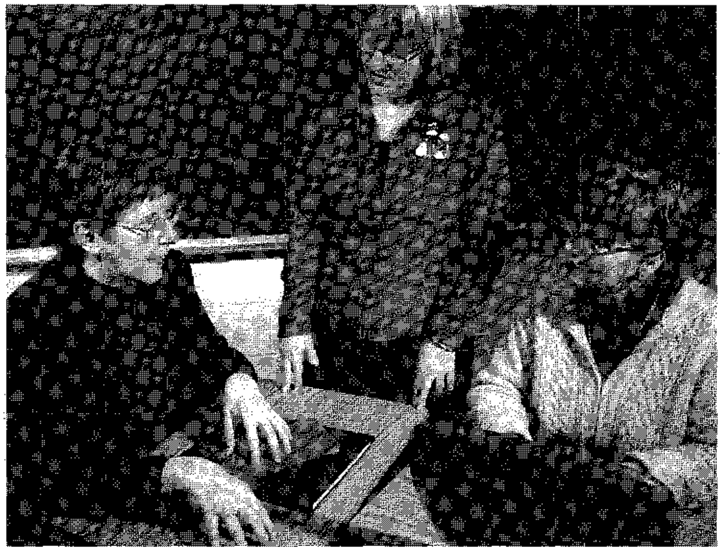
Why License Teachers?

Requiring teachers to be licensed seems intuitively appropriate, given their role in society. Society requires many professionals who work directly with the public, including lawyers, psychologists, nurses, and doctors, to be licensed in order to protect the public. These regulations ensure that the professionals working in these fields have the skills and knowledge required to perform their services effectively and ethically. Should the professionals who spend six to eight hours alone in a room with 20 vulnerable children be subject to any less scrutiny? Licensing teachers is a way to assure the public that teachers are competent, are qualified, and will, at the very least, do no harm.

The most basic purpose of licensure is to give the state control over who can teach, preventing those convicted of sexual misconduct, child abuse, or other relevant offenses from becoming or continuing to be school employees. Most states require a criminal background check as a part of licensure. Similarly, states screen with tests of basic skills in literacy and

by MARY E. DIEZ

mathematics to ensure that the academic skills of would-be teachers are at least above some minimum threshold.



The problem with many states' licensure requirements has been that they have not been thoughtfully developed as a coherent picture of what teachers need to know in order to be effective.

Like other bureaucratic processes, licensure also serves as a means of enforcing regulations. Most states have prescribed the coursework that candidates for licenses must take, down to the number of credits in specified courses with specified content. In Wisconsin, for example, elementary teachers cannot receive an initial license or renew a current license without proof of having completed coursework in how to teach reading using phonics. For an initial license, that proof comes as part of a 12-hour concentration in reading and language arts. In many states, teachers must earn additional professional development credits (usually six credits every five years) in order to renew their licenses, but teachers can earn these credits in areas that bear little relationship to their practice. Many teachers select courses based on their convenience and cost instead of their professional value. At a recent Wisconsin Education Association Council's convention, banners over the booth of one out-of-state professional development provider read, "Three credits—five days!" States have written their policies in such a way that the number of credits earned has become more important than the skills learned, both in teacher education and continuing professional development.

If you want to teach high-school social studies in Wisconsin, state law requires that you study the economics of dairy cooperatives.

The problem with many states' licensure requirements has been that they have not been thoughtfully developed as a coherent picture of what teachers need to know in order to be effective. Rather, they have been piecemeal collections of basic knowledge and skills, coupled with trendy political issues that have been voted into statute. For example, state legislators added the study of environmental education, human relations, and conflict resolution to the licensure requirements in Wisconsin. And if you want to be licensed to teach high-school social studies in Wisconsin, state law requires that you study the economics of dairy cooperatives.

Licensure, for the most part, has served its basic purpose of keeping dangerous people out of the classroom, ensuring that teachers are literate and numerate, and enforcing training and coursework requirements. Furthermore, in the past 15 years, the route to licensure has become less rigid and bureaucratic. In some states, dual systems of licensure—for "regular" and "alternative" routes—have emerged, sparking lively discussion among those who continue to be held to the "regular" requirements. Nevertheless, the process of licensing teachers has come under scrutiny in an increasing number of states, as critics ask whether current licensure practice actually leads to higher student achievement. More and more, the challenge to business as usual in teacher licensure is resulting in a complete overhaul of the process.

A Meaningful License

Reformers have asked three questions in seeking to develop a licensing system that guarantees not only that teachers have met a series of requirements, but also that they are prepared to enter classrooms as effective teachers. They concern:

- *Standards*: What should teachers know and be able to do in order to work effectively with learners?

- *Assessment*: What counts as evidence that teachers have learned these skills and knowledge? In other words, how should their performance be measured?

- *Training*: What are appropriate routes to developing these skills and knowledge?

The standards issue emerged in the mid-1980s, as states began to develop standards for what K–12 students ought to learn at each grade level. This had clear implications for the knowledge and skills that teachers who work with K–12 students need to possess. No less important was the emergence of alternative routes to teacher certification, such as the Troops to Teachers program for retired military personnel. Because

PHOTOGRAPHS COURTESY OF ALVERNO COLLEGE

The key disposition is respect for the individual learner, as shown in teachers' efforts to connect with their students and to tailor instruction to their students' needs.

licensure in most states had come to mean completing a series of mandated courses, the process wasn't equipped to handle candidates who brought important skills and experiences to the classroom—skills and experiences that would allow them to move more quickly into the classroom if it weren't for the state's bureaucratic requirements. At first, states grappled with what appeared to be inequitable systems—a prescribed system for regular candidates, and a loose system for career changers. Legislators and reformers have addressed this inequity by focusing on standards rather than on requirements. Indeed, the American Association of Colleges for Teacher Education has stated that while all teachers should be held to the same standards, alternative routes to meeting the standards can and must be developed.

The focus on licensing standards got a boost from the work of the National Board for Professional Teaching Standards, even though the National Board was concerned more with certifying accomplished teachers than with state licensure. The National Board's standards of good teaching practice guided the work of the Interstate New Teacher Assessment and Support Consortium (INTASC) as it began to develop a set of prototype standards for teacher licensure. At this writing, 37 states have used those standards in revising their licensure process.

While all teachers should be held to the same performance standards, alternative routes to meeting the standards can and must be developed.

The INTASC standards attempt to richly describe what the role of the teacher demands. They are outlined in three parts—knowledge, dispositions, and performances—making clear that teaching is a complex endeavor. While subject-area knowledge is privileged in the standards—INTASC has developed specific standards for science, mathematics, English language arts, and special education, and will soon release those for the elementary-school level, social studies, and foreign language—the standards also make clear that knowledge of child development, learning theory, and teaching approaches is essential. And some attention is being paid to candidates' attitudes toward their subject matter and toward learners in the statements of "dispositions" that are incorporated into each of the standards.

To illustrate, let's examine INTASC's Standard 3, which reads: "The teacher understands how students differ in their approaches to learning and creates instructional opportunities that are adapted to diverse learners." (See sidebar on pp. 12–13 for full text of the standard.) The knowledge required of teachers to meet this standard builds on the findings of cognitive psychology in stipulating that teachers must understand various learning styles and approaches to learning, as elucidated in Howard Gardner's work on multiple intelligences. It also requires that they know how to handle specific differences that matter in the classroom, like students for whom English is not their first language or students with special needs. Teachers must understand and adapt their instruction to their students' previous experiences, language, culture, and community values.

Dispositions identify the attitudes and values that guide and support the work of the teacher. In the case of Standard 3, the key disposition is respect for the individual learner, as shown in teachers' efforts to connect with their students and to tailor instruction to their students' individual needs. Finally, the performance standards describe the application of teachers' knowledge and disposition in the tasks they undertake in the classroom.

The standards issue, however, is not without controversy. Standards can be just as politicized as the old state codes often were. Special interests may attempt to hijack standards, and they sometimes succeed in using them to promote a particular philosophy or approach, as we saw in the controversy surrounding drafts of the K–12 history standards. Standards cannot, by themselves, force people to rethink teacher education. Indeed, some teacher educators resist moving from the old practice of course-based teacher education and simply overlay a new surface of standards language. The past has perhaps conditioned us to look at courses as separate "bits" that together count as a program. In a standards-based approach, what candidates do in courses contributes to their meeting standards, but the standards provide a framework that cuts across courses. At times, state bureaucrats and teacher educators seem to be attempting to turn the standards into the same kind of check-off approach that counting courses represented. I have visited some programs where one standard is reduced to a single project required in course A and another to a single paper in course B. Given the rich language of INTASC's Standard 3, a single project or paper could not possibly provide evidence of mastery.

Standards represent a major shift, from focusing on inputs to identifying key results. Standards invite all those involved in the endeavor—current teachers, school administrators, teacher educators, and state education officials—to a conversation about the meaning of good teaching and its impact on student learning. Wisconsin's revision of the public instruction code for teacher education is using the meaning of the standards to allow both regular and alternative teacher-education programs to design their own approaches in contrast to the "locked-in" course requirements of the past. Institutions and alternative providers alike must make a case for how they've adhered to the standards.

The first responsibility of the licensure process is thus to make clear, through the development of standards, what the license stands for—indeed, what it guarantees about the knowledge, skills, and dispositions of the teacher who holds it.

Assessment

The most critical issue for licensure is assessment. What will count as evidence that a candidate is ready to be licensed? What makes the guarantee of the license meaningful?

In the past, most states granted a license if the candidate completed the courses in the approved program of a college or university. Few would seriously argue that completing courses is automatically the same as developing the knowledge, skills, and dispositions required for effective teaching. Yet states and

The New Standards in Teacher Education

The standards of the Interstate New Teacher Assessment and Support Consortium (INTASC) outline the knowledge, dispositions, and performances necessary for effective teaching. Herewith, INTASC's Standard #3 and the knowledge, attitudes, and skills necessary to meet it.

INTASC Standard #3: The teacher understands how students differ in their approaches to learning and creates instructional opportunities that are adapted to diverse learners.

Knowledge

- The teacher understands and can identify differences in approaches to learning and performance, including different learning styles, multiple intelligences, and performance modes, and can design instruction that helps use students' strengths as the basis for growth.

- The teacher understands and can provide adaptations for areas of exceptionality in learning—including learning disabilities, visual and perceptual difficulties, and special physical or mental challenges.

- The teacher knows about the process of second-language acquisition and about strategies to support the learning of students whose first language is not English.

- The teacher understands how students' learning is influenced by individual experiences, talents, and prior learning, as well as language, culture, family, and community values.

- The teacher has a well-grounded framework for understanding cultural and community diversity and knows how to learn about and incorporate students' experiences, cultures, and community resources into instruction.

Dispositions

- The teacher believes that all children can learn at high levels and persists in helping all children achieve success.

- The teacher appreciates and values human diversity, shows respect for students' varied talents and perspectives, and is committed to the pursuit of "individually configured excellence."

- The teacher respects students as individuals with differing personal and family backgrounds and various skills, talents, and interests.

- The teacher is sensitive to community and cultural norms.

- The teacher makes students feel valued for their potential as people, and helps them learn to value each other.

Performances

- The teacher identifies and designs instruction appropriate to students' stages of development, learning styles, strengths, and needs.

teacher educators alike have accepted—and some continue to accept—the proxy as appropriate.

On many campus visits I've made as part of my work as a member of the board of examiners for the National Council for the Accreditation of Teacher Education, I've asked faculty members about the relationships between the statements in the front of their syllabi and the process of determining grades, outlined at the end of the syllabus. In these syllabi, statements of goals regarding knowledge, skills, and disposition are usually written well. The key is whether students can pass the course without demonstrating that they have met the goals. Rarely are assessments clearly linked to the goals. Too often what passes for assessment is less focused on quality of performance than on issues of format and punctuality—both of which can be important, but are not central. Point systems may give a nod to quality, but they don't often make explicit the criteria for quality. And faculty members often admit that students can earn enough points to pass without demonstrating the aspects of standards intended to be developed in the course.

- The teacher uses teaching approaches that are sensitive to the multiple experiences of learners and that address different learning and performance modes.
- The teacher makes appropriate provisions (in terms of time and circumstances for work, tasks assigned, communication and response modes) for individual students who have particular learning differences or needs.
- The teacher can identify when and how to access appropriate services or resources to meet exceptional learning needs.
- The teacher can identify when and how to access appropriate resources to meet the needs of students with particular talents.
- The teacher seeks to understand students' families, cultures, and communities, and uses this information as a basis for connecting instruction to students' experiences (e.g., drawing explicit connections between subject matter and community matters, making assignments that can be related to students' experiences and cultures).
- The teacher brings multiple perspectives to the discussion of subject matter, including attention to students' personal, family, and community experiences and cultural norms.

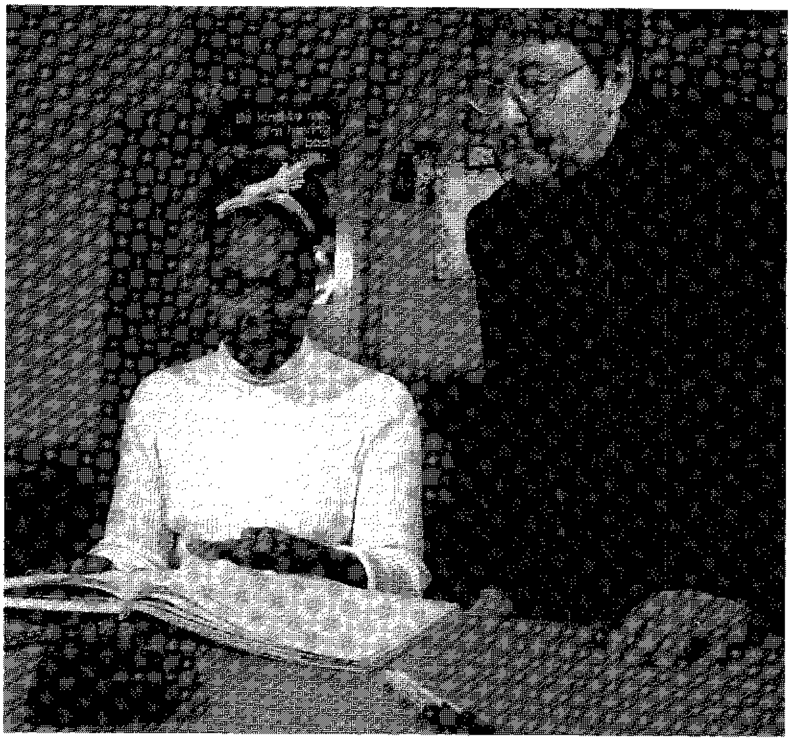
Assessing whether a candidate has adequately demonstrated an aspect of a standard first requires clear criteria for what would count as such a demonstration. For example, in an Alverno College assessment focused on teaching writing in elementary school, candidates work with samples of actual student papers from a local district. After working with the district's rubric, each candidate is given a sample paper to assess; the task involves first providing the evidence to support the candidate's judgment about the developmental level of the 5th grader's performance. Then, building on that judgment, the candidate develops an instructional plan based on her diagnosis of the learner's strengths and needs.

> In Oregon, teacher candidates put together work samples demonstrating the links between teaching and K–12 student performance.

Criteria for successful completion of this assessment include accurate use of the rubric (compared with the district teacher's ranking of the paper) and effective reasoning based on specific aspects of the student's writing. For the second part of the task—developing an instructional plan—criteria include the designation of two to three appropriate areas that the teacher would work on next with this learner and a clear rationale for the links between the student's needs and the instructional plan.

Some programs simply ask teaching candidates to provide samples of different types of lesson plans, unconnected to school context, as artifacts that go into a portfolio to demonstrate meeting the INTASC standards. Such sample plans, developed in the absence of real learners' needs, do not provide clear evidence that a candidate can "create instructional opportunities that are adapted to diverse learners."

I've argued that a person could conceivably demonstrate some, most, or even all of the standards we require with little or no input from a teacher-education program. Shouldn't we recognize, for example, the understanding of developmental psychology that literate parents have developed through the experiential learning involved in raising a child? And if those parents can transfer the understanding gained through experiential learning to designing developmentally appropriate instruction for learners, shouldn't we be able to validate that knowledge and skill? If we took assessment seriously, we could develop learning experiences and assessment processes to fit the accumulated knowledge of candidates—and not have everyone move lockstep through the same series of courses.



During student teaching, candidates put together a portfolio that illustrates their work with K–12 learners, assessing their work and giving them feedback and showing how their planning addresses the needs of their class.

A number of significant problems with assessment as related to licensure continue to dog the credibility of the process. Most important, while the standards describe teaching as a set of highly complex tasks, much of what currently passes for teacher testing follows a reductionist model—looking not so much at what's important but at what's easy to measure. Critical aspects of the standards are ignored because they are difficult to assess. For example, for Standard 3, we want to know how the teacher works with a group of 25 students, with a range of needs, and adapts instruction to meet their individual needs. Assessing that kind of instruction is an expensive, complex proposition. It requires documenting what the teacher knows about the learners and how she uses that knowledge in day-to-day work with the range of needs. Less expensive might be a multiple-choice test in which a teacher identifies the names of theorists and links them to their theories, but this does not begin to provide the same level of information. In the current world of teacher licensing, cost is an issue, and so many states let multiple-choice tests stand as proxies for the more complex assessment that would provide evidence of whether candidates meet the standards.

Some candidates could prepare outside of a teacher-education program and still be able to demonstrate the necessary skills. Some, but not many.

In order to make licensure a mark of quality, assessment needs to document knowledge and performance in practice and over time. In the new Wisconsin requirements, teacher-preparing institutions and alternative providers alike must give evidence of multiple, complex assessments tied to the standards for the initial license. At Alverno, we embed complex assessments in both liberal arts and teacher-education curricula. For example, our students interact with and observe four-year-olds to see how developmental theory is played out in the way children approach the world. They develop and teach lessons in classrooms over five semesters, initially tutoring one or two students and then teaching small groups and whole classes. They learn the expectations of a local district's science curriculum and not only show the ability to assess 6th grade science projects using the district's rubric, but also plan the next steps in designing instruction to meet the students' needs. Faculty members provide feedback on these assessments, assisting teacher-education candidates as they continue to build their knowledge and skills in preparation for their role with learners. During student teaching, candidates put together a portfolio that illustrates their work with K–12 learners, assessing their work and giving them feedback and showing how their planning addresses the needs of their class. Thus the portfolio includes videotapes of the teacher working with students across a range of lessons, lesson plans, student work samples with teacher feedback, and the teachers' continuing reflections on why they are doing what they do.

Similarly, in Oregon, teacher candidates put together work samples demonstrating the links between teaching and K–12 student performance; they do this during student teaching and again in the first year of practice. A work sample usually includes a description of the class's make-up, with demographic and other data, lesson plans, data from pre-tests and post-tests, and analysis of the learning gains made by students. In Connecticut, the BEST program requires new teachers to put together a portfolio of their performance during their second year that demonstrates their planning, instruction, assessment, and feedback practices. Continued licensure is contingent on meeting the criteria for adequate performance.

The Connecticut BEST portfolio assessment is more expensive than a multiple-choice test, but the evidence it provides is far superior. Teachers who serve as assessors receive rigorous training in the observation and assessment of teacher performance against well-articulated standards. This training of teachers as assessors for the process is also helping to build a different culture of learning, based on the statewide conversation about what constitutes good teaching.

The second responsibility of the licensure process is thus to ensure, through valid and meaningful assessment processes, that the teacher can do what the job requires.

Training

James W. Fraser argues that it's time to decouple the relationship between teacher education and licensure. I agree that the status quo, in which the state blesses college- and university-based teacher-education programs but never seeks evidence of teachers' performance in real classrooms, must be changed. Just taking courses does not ensure that candidates become quality teachers, just as knowing a subject doesn't guarantee an ability to teach it. The shift toward standards and meaningful assessments opens up more possibilities. Now, whatever route candidates take to licensure, they must demonstrate the knowledge, dispositions, and performances outlined in the standards. Some candidates could prepare outside of a program and still be able to demonstrate the necessary skills—some, but not many.

Two arguments support maintaining a connection between state requirements for licensure and the programs that prepare teachers to stand for licensure, whether those programs are housed in higher-education institutions, in school districts, in other organizations, or in collaboratives involving any combination of groups. First, such a connection can protect the public by identifying programs that have provided evidence of offering legitimate, credible preparation that is linked to the standards to which new teachers will be held. In the past, colleges and universities were assumed to have a lock on legitimacy; I would instead like to see *all* providers produce evidence of their credibility. Not everyone who wants to put out a shingle to prepare teachers is necessarily qualified to do so.

Second, such a connection can protect the candidates who are seeking a viable route into the classroom by ensuring that programs provide candidates with the opportunity to learn, through both meaningful experiences and effective developmental assessment. Assuming for the moment that some candidates could stand for assessment and be licensed outside of a formal program, I believe that those who choose to go through a program—again, wherever it is provided—should have some assurance that what they get is worth the money they pay for it.

Opening up the range of providers might encourage more cutting-edge approaches. With advances in both assessment and technology, providers of teacher preparation ought to be able to design an almost infinite array of paths to demonstrating performance that meets the standards.

Thus, where Fraser argues for no relationship, I would urge a partnership in which the state encourages the responsible development of a range of program options. In fact, Alverno is involved in two such alternative programs, one with an urban school district and two other colleges, the other with three foundations, a group of independent schools, and another college.

We need to avoid the narrow thinking that says there's one best way to prepare teachers. A few years ago, the Holmes Group argued the necessity of five- and six-year programs. However, the Federal Awards for Excellence in Teacher Preparation, which sought to find programs that have documented impact on K–12 student learning, honored three undergraduate programs (Alverno, Samford, and Eastern Carolina) and one graduate-level program (Fordham) in 2000. Many approaches may lead to the outcomes that we seek; states should not fall into the trap of specifying only one or a few approaches.

Thus it is the third responsibility of the licensure process to sanction approved routes and programs to obtaining the license, wherever they are established.

The Case for Licensure

Thoughtful critics have suggested that teacher licensure is an unnecessary burden when it is placed on candidates with appropriate content majors and related work experience. Underlying this argument is the notion that knowledge of student development and modes of learning either is best learned on the job or is not relevant.

While it's hard to make the case for licensure when it requires only taking a collection of courses, it's even harder to argue against a standards-based licensure system that ensures that every child will have a well-trained, high-quality teacher. Such a system depends on complex processes of assessment that incorporate basic skills, content knowledge, pedagogical knowledge, and performance in a school setting with real students.

Perhaps another Wisconsin analogy will help focus the case for meaningful licensure. In Wisconsin, a licensed veterinarian has to pass three types of exams: a written standardized test, an oral interview, and a demonstration of surgery on a small animal. Should the requirements be any less stringent for those to whom we entrust our children?

–Mary E. Diez is a professor of education and graduate dean at Alverno College in Milwaukee, Wisconsin.

a TENUOUS HOLD

Education schools have lost the confidence of the public and policymakers alike. They'll need to relinquish their monopoly on teacher preparation in order to gain it back

As dean of an education school, more and more I find myself asking, to my own surprise: Is it time to sever the link between the university programs that prepare teachers and the public bodies that certify them? For half a century, college- and university-based teacher-education programs have held a monopoly position in operating, in partnership with state governments, virtually the only route to the teaching profession in public schools. This has guaranteed us a fairly constant stream of students, revenues, and state and federal funding. These benefits have come with strings attached, of course: states have mandated much of the curriculum that prospective teachers must take.

The question is, has our monopoly position and the states' intervention helped or undermined the quality of our programs? Has our monopoly made us unresponsive to the needs of our students and the schools that hire them? Have state mandates distracted us from designing a coherent, innovative, and high-quality curriculum? I think so. As the education system makes a transition to a focus on results-based accountability combined with more flexibility in school design, it's time for those of us in the teacher-education profession to become more results-based and flexible too. We ought to concentrate on offering the best possible professional training programs while leaving it to others—state and local governments and individual schools—to decide who is qualified to be certified and hired. The breaking up of our monopoly would force us to convince students, their tuition-paying parents, and the school districts that do the hiring that our programs

PHOTOGRAPH BY MIKE DOBEL/MASTERFILE

BY JAMES W. FRASER

The reality is that we haven't gained the respect of our colleagues in higher education, and many of our graduates claim to have learned little of any value in education school.

produce teaching candidates who are more qualified and skilled than candidates who obtained their training elsewhere or who come in with no training.

Policymakers and the public have realized that one of the keys, if not *the* key, to raising student achievement across the board is to improve the quality of the nation's teachers. To my mind, nothing is more important to school quality than placing a passionate, dedicated, knowledgeable, and highly skilled teacher in every classroom. Doubts about the academic abilities of some of the nation's teachers—as measured by various tests of student learning—have led to doubts about the university programs that prepare teachers. The conservative backlash against progressive education has also planted seeds of doubt about the content of the education provided in education schools. Moreover, the national teacher shortage has forced policymakers to create alternative pathways to teaching in order to encourage mid-career changers and talented college graduates who didn't complete education programs to go into teaching. The success of programs like Teach for America and other alternative-certification programs has convinced many that a teacher needn't attend an education school in order to be effective. Education schools need to recognize these doubts and answer them effectively.

This is not a call to move the professional preparation of teachers and administrators from the academy to the "real world" of practice in the schools. Such a move was a major part

"The reality remains," says Northeastern University School of Education dean James W. Fraser, "that we haven't gained the respect of our colleagues in higher education, and many of our graduates themselves claim to have learned little of any value in education school."

PHOTOGRAPHS BY MARK ALCAREZ

of the conservative agenda of British prime minister Margaret Thatcher, who sought to denigrate academic learning of any kind. A similar anti-intellectualism has come quickly to the United States. Classroom experience is a crucial part of any teacher's training, but it is only a part. The apprenticeship approach works well in preparing people for a profession as it is currently practiced. It does not work very well in preparing people to be reflective or to innovate and be leaders of change. And if today's schools need anything, it is some pretty basic reflection, rethinking, and reconsidering.

Nevertheless, we in teacher education do not have a stellar record of producing change agents, and we have failed at the task of policing ourselves. There are many wonderful teacher-preparation programs today, many more than there were only a few years ago. Leaders in teacher education have succeeded in raising the academic standards of their programs and promoting more interaction between education faculty and arts and sciences faculty. We have also answered the criticism that education programs were long on theory and short on practice by linking our programs much more effectively to schools through yearlong internships and through the development of a clinical faculty of school-based teacher educators who are truly part of the teacher-preparation team. But the world of teacher education remains uneven. The teacher educator who has not been in a school in years; the one who advocates "multiple modes of instruction" but only lectures in class; the one who is just plain boring: these people are still all too real. Above all, the reality remains that we haven't gained the respect of our colleagues in higher education, and many of our graduates themselves claim to have learned little of any value in education school. This alone should cause us to do some dramatic rethinking.

A Brief History

The current university-state partnership in preparing and then certifying teachers is relatively new. For much of the nation's history, school districts often ran their own teacher-preparation programs. In many cases, elementary-school teachers were simply graduates of the local high school. Normal schools, which began in the 1830s, were in their early years much closer to being high schools than the equivalent of colleges. Most expanded their curriculum for future teachers to the equivalent of a baccalaureate degree only in the early decades of the 20th century. For all the debates over the virtues and vices of progressive education, before World War II the vast majority of Americans attended only elementary school, and they were taught by normal-school or high-school graduates who had often studied for one or two years at a school of pedagogy. Only in the mid–20th century did the notion that every teacher should have a college degree and some sort of professional preparation even begin to take hold. My home state of Massachusetts, long known as a leader in school reform, began requiring a bachelor's degree for teachers only in 1954. At about the same time, state departments of education began the process of monitoring college-based teacher-preparation programs using standards that were usually the result of careful negotiations among the state, the representatives of the college, and—in some cases—representatives of teachers and school administrators.

There is no question that these mid-century reforms made a significant difference in the quality of teaching across the country. In many states the old normal schools were transformed into a state college system after World War II. Higher education expanded to offer amazing new opportunities to the waves of GIs returning from the war and many others who had previously been excluded. And a good number of these college graduates—especially the female graduates—became teachers. In the 1980s many states raised the standards once again. Especially after the 1986 reports of the Carnegie Forum on Education and the Economy and the Holmes Group of education deans, most states abolished the old education majors and required that future teachers combine a rich and full baccalaureate program with significant preparation in pedagogy, either as an undergraduate minor or through a fifth year or other kind of graduate program. The teachers of the past half-century have been the best prepared of any generation of teachers in history. They have had better undergraduate educations and deeper knowledge of their subject matter and the complex process of teaching and learning than any who preceded them.

Still, the nation is not satisfied with the quality of the nation's teachers, and rightly so. Much of the blame has fallen on education schools. There are many other culprits, from the

Teacher Preparation at Northeastern University

At Northeastern, would-be teachers must complete a major in the College of Arts and Sciences in a field other than education as well as the following coursework.

- Introduction to Education
- Child Development, Learning, and Education
- Learning and Teaching Process
- Teaching Children's Literature and the Arts
- Integrating School Curriculum Through Social Studies
- Teaching Math and Science to Children
- Teaching the Language Arts
- Introduction to Special Education
- Health and Physical Education in Elementary School
- Four cooperative education (practicum) experiences
- One semester of student teaching

SOURCE: NORTHEASTERN UNIVERSITY

If we have high-stakes tests and all the public attention that goes with them, we should be able to trust school districts and schools to make wise decisions about whom to hire.

unequal funding of school districts, which leads to gross disparities in teacher pay, to the sheer lack of respect afforded to teaching as a profession. But we in teacher education are ultimately responsible for preparing prospective teachers to be great teachers. So as more and more states adopt various forms of high-stakes testing for both students and teachers, and as the reports of failure on a massive scale continue to proliferate, we in teacher education face harsh scrutiny. The logic is unavoidable: if the schools are failing, then the teachers are failing. If the teachers are failing, then the programs that prepare them must be failing too. Deeper analysis has raised many questions about the quality of some standardized tests. But high-stakes tests are here to stay. They speak for a public that has lost faith in the bright assurances of professional educators that teachers were well prepared and that students were learning.

At the same time, and arising from the same frustrations, many states have sought to impose even tighter regulations on the current state-approved college- and university-based teacher-preparation programs. In 1998 the U.S. Congress joined in this effort with its requirements for tough new reporting measures—the Title II Report Cards—that every teacher-preparation program in the country now files and that will be made public across the nation in the spring of 2002. Leaders within the teacher-education enterprise itself have also demanded much better preparation for future teachers. The 1996 report of the National Commission on Teaching & America's Future made a series of dramatic recommendations, including a call for states and school districts to "get serious about standards, for both students and teachers" and for a thorough restructuring of teacher preparation, recruitment, and professional development. Meanwhile other reformers wish to dismantle the entire certification system and afford public school principals the same privilege as private school principals: the ability to choose any teacher, certified or not, based on a personal assessment of their skills and attitude.

A Modest Proposal

Overhauling the system of teacher education first requires that education schools recognize that for many in the larger public—the parents of today's schoolchildren, the taxpayers who foot the bill, and the elected officials who represent them—the current system of teacher preparation is seen as part of the problem, not the solution. While we may complain that such judgments are unfair, we need to address the reality that, fairly or unfairly, teacher education has lost its legitimacy in many circles.

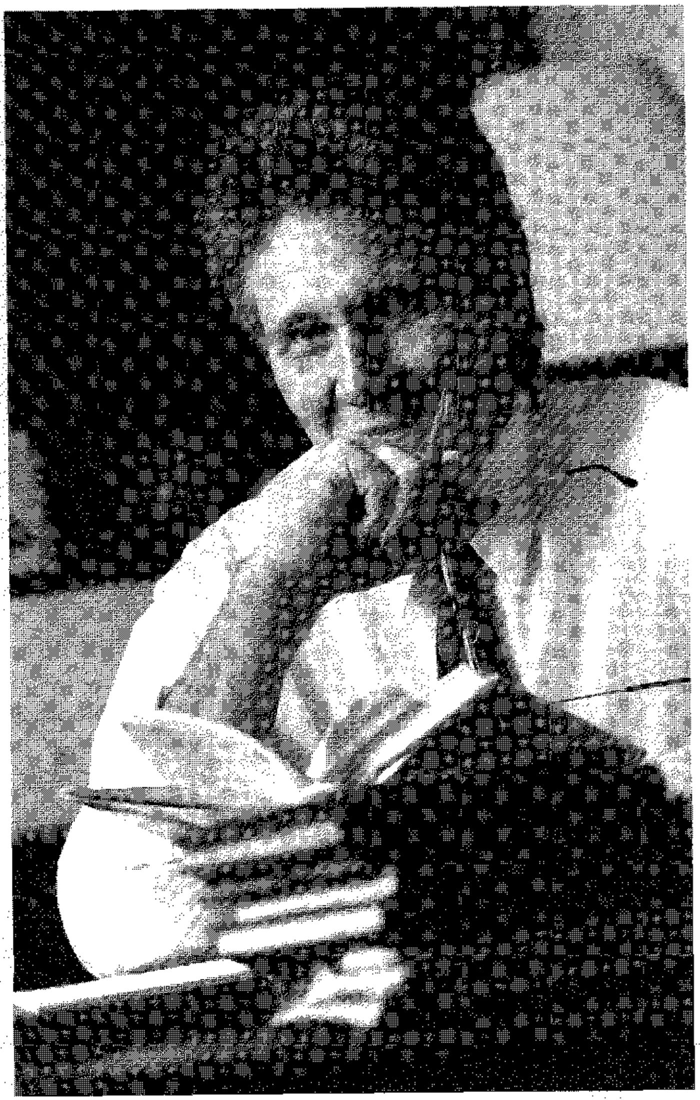
As someone who has devoted a large part of my professional career to teacher preparation, this situation worries me. I am not a fan of the old system, but I am deeply worried that "alternative routes" are going to create a generation of teachers who do not have the fundamental skills, knowledge, and inclinations needed to support student learning. I fear models that reinforce the status quo in our schools, because for me the status quo in American education is unacceptable. Too many kids are left out, marginalized, or taught that their failures are their own fault. Too many kids are excluded from the benefits of today's economy and from participation in tomorrow's democracy. What is to be done?

Perhaps the time has come for those of us in higher education to simply step out of the teacher-certification business. We should let school districts and, ideally, in many cases, individual schools make the certification and the hiring decisions. States, in turn, should step out of their role in regulating university-based teacher-preparation programs. Both the schools and the higher-education programs will thrive, I believe, in such a new atmosphere of freedom and accountability.

The role of university-based teacher preparation should be different—and if we do the preparation right, more significant. Let the schools and school districts certify and hire whom they will. Our role as teacher educators will be to provide programs that add such clear and obvious value that school districts prefer our candidates to others. If we are doing a good job, our graduates will obviously be the best teachers. If they are not, we will have only ourselves to blame.

If we give up the monopoly that has protected us in the past, a monopoly that is gone whether we like it or not, we should be given much more freedom to design the kinds of programs that we believe will truly prepare the best teachers. If the state no longer certifies teachers because they are our graduates, the state also must relinquish any right to regulate our curriculum. Let us clearly separate the two functions. We would not offer any courses "because the state mandated them." We would be free to design the best curriculum we could imagine, to seek accreditation or not, as it served the education of our students and—more important—their students. Let our quality be judged by our graduates' ability to get the best teaching jobs and by the learning of students in their classrooms.

I can already hear the first objection: "But what about nepotism? What about a politicized hiring process?" These are not idle worries. In the much romanticized "good old days"



"I cannot imagine quality teacher education that is not based on a rigorous academic curriculum alongside a series of increasingly complex and extended field experiences," says Fraser.

before teachers were certified, many a school board member hired teachers based on who could offer the best pay-off, in cash or in promised votes. Many a school administrator hired cousins and friends, sometimes their not-too-bright cousins and friends, who could not find other employment. The schools were the great employment agency of last resort, not only in corrupt big cities, but also much more in small towns, where fewer outside eyes were watching. But society has undergone a dramatic change. It is not that 21st-century America is magically free of corruption and cronyism; it is that now there are no places where outside eyes are not watching.

Ironically, it is the high-stakes testing movement that has made nepotism a losing proposition and the need for certification less pressing. In an accountability system where the very livelihood of educators depends on their schools' performance, what school board member, superintendent, or principal would hire an underqualified teacher and allow that teacher to bring down the school's or the district's test scores, thereby incurring the wrath of parents, state legislators, and the media? In the past it was possible to have Lake Wobegon schools where school leaders assured parents and the community that wonderful learning was happening and all of the children were above average. Today the averages are out there for everyone to calculate. With high-stakes testing, the stakes have simply become too high for anyone to risk his or her career by failing to exercise informed judgment about the quality of the teachers being hired.

If we have high-stakes tests and all the public attention that goes with them, we should be able to trust school districts and schools to make wise decisions about whom to hire. Those of us in higher education should design admission standards, curricula, clinical experiences, and exit examinations that will allow us to be sure that our graduates really are the best women and men for the jobs. If we do that, and if we succeed, our programs will flourish and our graduates will be able to compete quite successfully for jobs in the most wonderful of careers—teaching our nation's schoolchildren.

What will such programs look like? My guess is that the most successful of the unregulated programs will adopt many of the recommendations of the National Commission on Teaching & America's Future. I cannot imagine quality teacher education that is not based on a rigorous academic curriculum alongside a series of increasingly complex and extended field experiences. These experiences allow students to mix the very best teaching with thoughtful and intellectually rigorous reflection on those experiences, along with a high-quality academic curriculum that is rich in content knowledge at both the undergraduate and the graduate levels.

Those of us who live our professional lives in college- and university-based teacher-preparation programs must face a radically changed and changing reality. We cannot return to the old system of states' closely regulating our programs while guaranteeing ready access to the profession for our graduates by making us the only route to certification. It was not that good a system, and in any case it is gone. Our choice is to be defensive and to mourn for the past or to strike out in a new direction, trying something new and adventurous and engaging, for our students and ourselves. Who knows, it might even be more intellectually and professionally rewarding for us and—what really matters—it might begin to prepare a generation of teachers who can have the intellectual resources, the professional skills, and the moral courage to design schools in which no one is excluded, in which every child succeeds and thrives.

–James W. Fraser is dean of the school of education at Northeastern University in Boston, Massachusetts.

BREAK the LINK

The fact that schools of education could no longer rely on a captive body of aspiring teachers would expose them to the cleansing winds of competition.

Picture Gerard, a 28-year-old business consultant who majored in economics at Williams College and graduated with a 3.7 GPA. Gerard has been working for a consulting firm in Stamford, Connecticut, but is looking for a new, more fulfilling position. He has demonstrated strong interpersonal skills and work habits. In addition, though he didn't major in math, he aced several calculus courses in college. Yet if Gerard were to apply through normal channels to teach math at a junior high school in the Hartford public school system, his application wouldn't even be considered. Why? Because he isn't a certified teacher.

Why shouldn't a principal or a faculty hiring committee in the Hartford schools even be allowed to look at Gerard's application, to judge his qualifications against those of other candidates? The assumption undergirding the contemporary approach to teacher certification is that public school hiring personnel are either unable or unwilling to gauge the quality of applicants. Our response has been to embrace a bureaucratic solution that handcuffs the capable and incapable alike and supposedly keeps weak teachers out of the classroom. As a result, having discouraged or turned away Gerard and hundreds like him, many large school systems resort to last-minute fill-ins who teach on emergency certificates.

This is not to suggest, even for a moment, that candidates with "real world" experience or high GPAs are necessarily qualified or equipped to become

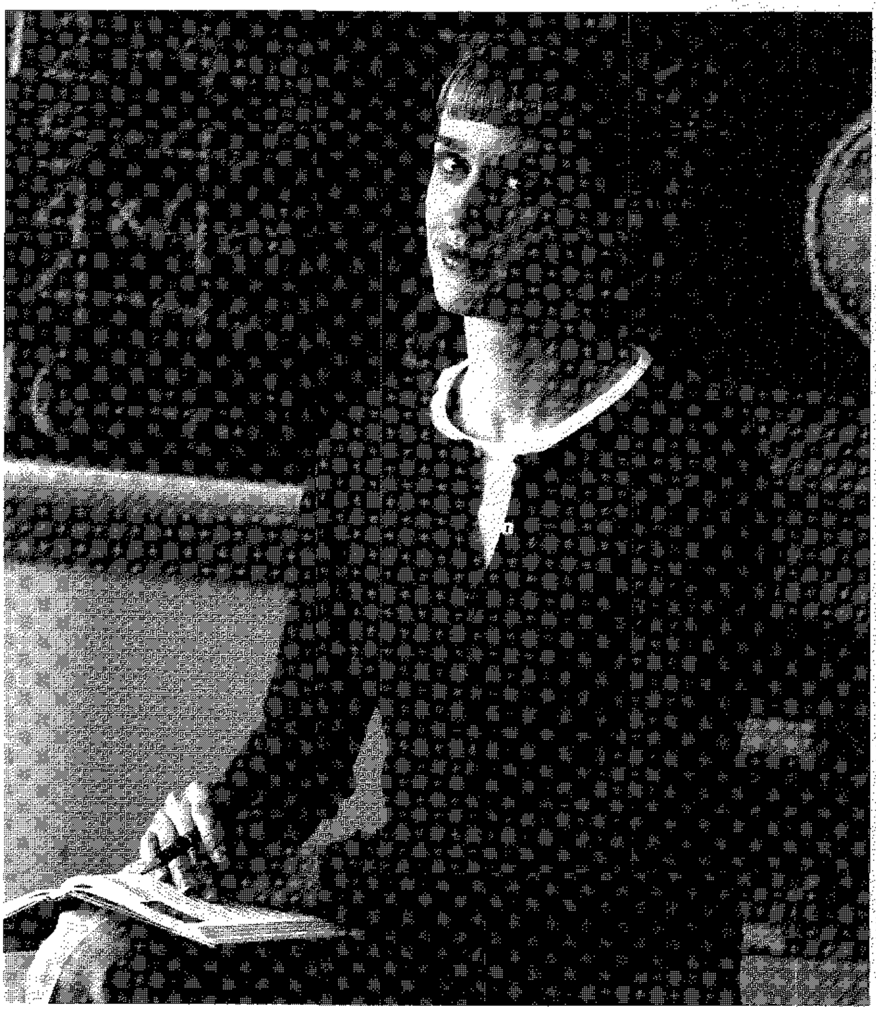
PHOTOGRAPH BY NORA GOOD/MASTERFILE

by FREDERICK M. HESS

teachers or that professional preparation for teachers is unimportant. It is only to say that some potential applicants *might* be more effective teachers than the alternatives that are currently available to public schools.

The central premise underlying teacher certification is that—no matter what their qualifications are—anyone who has not completed the specified training is unsuited to enter a classroom and must be prohibited from applying for a job. Presumably, the danger is that, in a moment of weakness, a school official otherwise will mistakenly hire such an applicant rather than an appropriately trained teacher. It is essential to remember what we often seem to forget, which is that allowing someone to apply for a job is not the same as guaranteeing him employment. Making applicants eligible for a position simply permits an employer to hire them *in the event that they are deemed superior to the existing alternatives*. The argument against certification is not that unconventional applicants *will* be good teachers; it is only that they *might* be. If one believes this, case-by-case judgments are clearly more appropriate than an inflexible bureaucratic rule.

Imagine if colleges and universities refused to hire anyone who lacked a Ph.D. They would lose the talents and insights of "lay practitioners" like poet Maya Angelou, journalist William Raspberry, or former public officials such as Alan Simpson, Julian Bond, and Al Gore. The artists and writers "in residence" at dozens of public universities would fail to meet the criteria implicit in the public school certification model. Do we really believe that these universities are ill-serving their students by hiring people whom the public schools would consider unqualified?



What is needed is a competitive certification process that establishes key criteria for entry into the teaching profession; gives public schools greater freedom to hire and fire teachers; and treats teachers like professionals.

PHOTOGRAPH BY PIERRE TREMBLAY/MASTERFILE

Competitive Certification

The theory behind certifying or licensing public school teachers is that this process elevates the profession by ensuring that aspiring teachers master a well-documented and broadly accepted body of knowledge and skills important to teaching. Supporters of teacher certification often make analogies to professions like law and medicine, where being an effective professional requires the acquisition of vast knowledge and skills. Licensure in these professions ensures at least minimal competency and boosts the public's confidence in members of the profession.

The problem is that no comparable body of knowledge and skills exists in teaching. Debate rages over the merits of various pedagogical strategies, and even teacher educators and certification proponents have a hard time defining a clear set of concrete skills that makes for a good teacher. Yet most aspiring teachers are still forced to run a gauntlet of courses, requirements, and procedures created by accredited training programs that vary dramatically in quality.

This is not to deny that teacher education can provide valuable training. After all, one may think that journalism schools produce better journalists without requiring all journalists to complete a mandatory set of courses before seeking work in the profession. Instead, it is assumed that a candidate's training is factored into the hiring process, along with considerations like aptitude, diligence, and energy.

Clearly some sort of screening process for aspiring teachers is essential; parents and the public rightly expect safeguards for those working with youngsters. What is needed is a competitive certification process that establishes key criteria for entry into the teaching profession; gives public schools greater freedom to hire and fire teachers; and treats teachers like professionals and their schools like professional institutions by allowing them to tailor professional development to meet the needs of teachers. Under such a model, aspiring teachers ought to be able to *apply* for a teaching job if they:

- Possess a B.A. or B.S. degree from a recognized college or university.

- Pass a test that demonstrates competency in knowledge or skills essential to what they seek to teach. The definition of "essential" knowledge or skills is obviously a loose one that can be interpreted in myriad ways and rightly should be different for those wishing to teach younger children or older students. The key point is to demand that teachers at least have an appropriate academic knowledge of the material they will be teaching.

- Pass a rigorous criminal background check. States conduct such checks now, but they tend to be compromised by the state's need to engage simultaneously in related certification paperwork.

Beyond these minimal qualifications, the competitive approach presumes that preparation and training are not only desirable but also essential, as is true in other professions where subtle skills and interpersonal dynamics are essential to effective performance. The questions are where to obtain this training and who should pay for it. Contemporary teacher preparation imposes nearly all of the costs on candidates by forcing them into a system of training that removes key incentives for quality and relevance in teacher preparation. The competitive model instead treats teachers as autonomous professionals able to make informed decisions about developing their skills and expertise. In short, the competitive model would substitute meaningful professional development for what is essentially a guild system funded by levying a significant tuition-based tax on aspiring teachers before permitting them to enter the profession.

The Assumptions of Certification

Over the years, an array of studies has sought to determine whether certified teachers serve students more effectively than uncertified teachers. There are two problems with this line of work. First, the methodological wrangling has often obscured the larger questions and the central assumptions of the certification model. Second, the case for certification is thin *whether or not* certified teachers boost student achievement more than their uncertified peers. The issue is *not* whether teacher education improves the performance of graduates, but whether we ought to—as best we are able—bar from teaching those who have not completed an approved preparatory program. Certification systems deny school administrators the ability to take their context or the promise of a particular applicant into account when hiring. Even if certified teachers are generally more effective than uncertified teachers, such a policy only makes sense if we believe that uncertified applicants are uniformly incompetent to teach or that school administrators cannot be trusted to assess their competence.

The allure of certification rests on three implicit assumptions. They are the beliefs that: 1) the training one receives while getting certified is so useful that the uncertified will be relatively ill-prepared; 2) certification weeds out unsuitable candidates; and 3) certification makes teaching more "professional" and therefore a more attractive career. However, each of these presumptions is problematic in the case of teacher certification.

As a general principle, certification is most effective when the licensing body ensures that aspiring professionals have mastered essential skills or knowledge and denies a license to inadequate performers. Licensure is most essential when a professional's tasks are critical and when clients may have trouble assessing a provider's qualifications. For instance, licensure is considered particularly appropriate for engineers, doctors, and attorneys because those who design bridges, tend us when we are ill, or defend our rights all perform tasks essential to our well-being and are frequently charged with aiding us at our most vulnerable. Moreover, it can be difficult for members of the public to know whether a bridge is properly designed, whether a doctor is performing appropriately, or whether an attorney is knowledgeable in the law. Licensing is not an assurance that these professionals are talented practitioners, but it does ensure that they have demonstrated an established degree of professional knowledge.

Educators are also charged with a crucial task. However, the oversight challenge is very different in education, where we have not established a specific, measurable body of skills or knowledge that teachers must master. Educational "experts" themselves argue that teaching is so complex that it can be difficult to judge a good teacher outside of a specific classroom context. This makes it difficult, if not impossible, to determine abstractly which aspirants possess satisfactory "teaching skills." Meanwhile, there is widespread agreement that colleagues, supervisors, and families have at least a proximate ability to gauge whether a teacher is effective. Given these circumstances, it is unclear how standardized licensing helps to safeguard teacher quality.

Such a conclusion does not require refuting the claims of teacher educators or the supporters of certification. It actually follows if one simply accepts their claims. Professional educators themselves have thus far been unable to explain in any con-

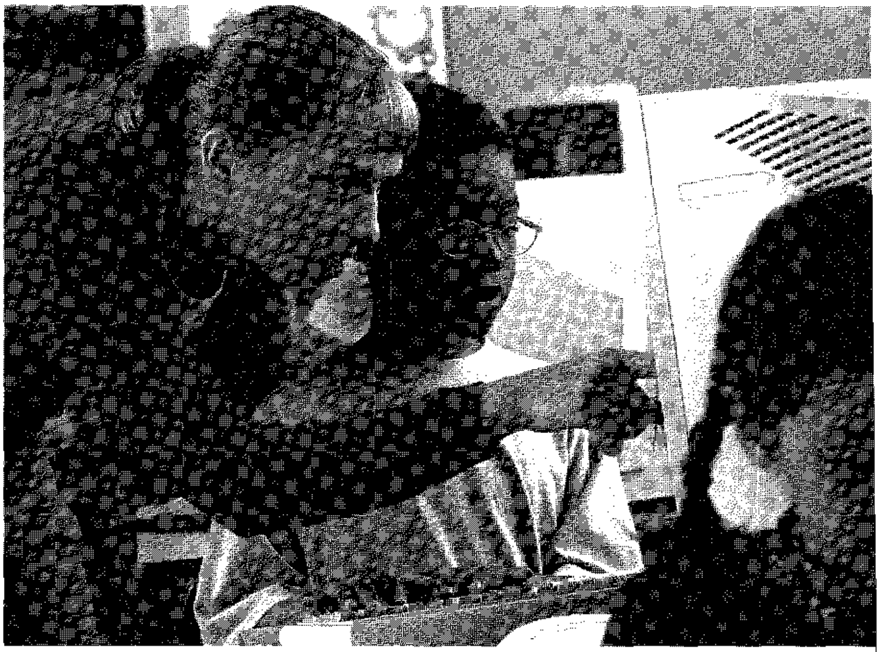
> **Aspiring teachers ought to be able to apply for a teaching job if they have a college degree and can pass a test of subject-matter knowledge.**

crete sense what makes a teacher competent or what teachers need to know and be able to do.

Consider the widely praised standards that the National Board for Professional Teaching Standards (NBPTS) has painstakingly constructed in 27 distinct fields, standards that certification's proponents have hailed as a breakthrough in quality control. The area where the NBPTS ought to have the easiest time creating straightforward standards is high-school math and science teaching, where there is widespread consensus as to what teachers are supposed to do. Even in these areas, however, the NBPTS's "exemplary" standards are so broad and vague as to make concrete judgments of competence nearly impossible. For instance, to receive National Board certification to teach high-school math, teachers are to demonstrate mastery of 11 standards, including: commitment to students and their learning, the art of teaching, reflection and growth, and reasoning and thinking mathematically. The board tries to clarify these standards by explaining, for instance, that "commitment" is interpreted as meaning that "accomplished mathematics teachers value and acknowledge the individuality and worth of each student, believe that all students can learn," and so on. Mastering the "art of teaching" is taken to mean that teachers "stimulate and facilitate student learning by using a wide range of formats and procedures." While these are certainly admirable sentiments, nowhere in the National Board's rarified standards is it clear how we are to gauge just what constitutes "competence" in these tasks. The result, unsurprisingly, is that the board has been assailed for the capricious way in which the standards are being interpreted and applied. Despite the best of intentions in the drafting of the INTASC standards, which Mary Diez discusses, a cursory read makes clear that they are plagued by the same ambiguities evident in the NBPTS standards.

For another prominent example, consider education professors Gerald Grant and Christine E. Murray's award-winning 1999 Harvard University Press book, *Teaching in America*. They identify five "essential [teaching] acts" that can be analyzed and taught: listening with care; motivating the student; modeling caring by hearing and responding to the pain of others, and by creating a sense of security in their classrooms; evaluating by clarifying, coaching, advising, and deciding on an appropriate challenge for this boy or that girl; and reflecting and renewing. How one is to teach these five "essential acts," much less determine whether a teacher has satisfactorily mastered them, are questions that Grant and Murray never address.

If clear standards of professional competence do not exist, we typically (and appropriately) hesitate to prohibit some individuals from practicing a profession. This is not to say that we think incompetence is acceptable in such a profession—only that we recognize licensing as an ineffective and potentially pernicious way to control quality. While licensure could protect community members (including children) from exposure to "bad" entrepreneurs or journalists, we do not prohibit some people from seeking to start businesses or work for a newspaper. Instead, we trust that potential investors or employers are the best judges of who ought to be supported or hired. If an aspiring writer or entrepreneur is unsuccessful, we trust that they will eventually be persuaded to find a line of work for which they are better suited. This free-flowing process fosters diversity and ensures that unconventional workers are given a chance to succeed.


The skills that teacher educators deem most important—listening, caring, motivating—are not susceptible to standardized quality control.

PHOTOGRAPH BY PIERRE TREMBLAY/MASTERFILE

Even in professions with clear knowledge- or performance-based benchmarks for certification, as in law or medicine, licensure is useful primarily as a way of establishing minimal competence. A medical or a law license is not imagined to ensure competence in ambiguous, subtle skills like comforting a patient or swaying a jury—skills analogous to the interpersonal relations thought crucial to teaching. Basing certification on such traits is difficult, because we may disagree about what they entail or how they can be assessed devoid of context. The skills that teacher educators deem most important—listening, caring, motivating—are not susceptible to standardized quality control. Emphasis on these qualities is the norm in professions like marketing, journalism, consulting, or policymaking, where a subtle blend of people skills, knowledge, and relevant expertise is required. In professions like these, where there are a number of ways for practitioners to excel but where it is difficult to know in advance how any particular practitioner will perform, the most sensible way to find talent is to allow aspirants to seek work and to permit employers to screen on a variety of criteria—such as education, experience, and references.

A Dubious Screen

While certification can serve to screen out aspirants who fail to meet a minimal performance standard, our current system is not designed to do so. Generally speaking, schools of education are not selective, flunk out few if any students for inadequate performance, and see that many of their teacher education graduates receive teacher licenses. The licensing exams are simple, and standards for passage are generally so low that the Education Trust concluded they exclude only the "weakest of the weak" from classrooms.

More than 1,300 institutions provide the training required for licensure. While defenders of the current approach to certification often focus on the certification programs at elite institutions, the top 25 education schools train less than 5 percent of the roughly 200,000 new graduates that teacher programs produce each year. It is the regional colleges, such as Illinois State University, Cal State–Hayward, and Southwest Texas State University—not the Stanfords and the Ohio States—that train and license the vast majority of teachers. The value of certification turns not on the quality of elite programs but on that of regional colleges.

Teacher-preparation programs neither screen out nor weed out weak candidates. Even at elite schools, such as UCLA or the University of North Carolina, where admissions rates are about 5 percent for medical school and 25 percent for law school, the M.Ed. programs (which include those seeking postgraduate training for teacher certification) accept more than half of their applicants. Moreover, education-school officials often make it clear that they do not see their mission as weeding out students during their course of study. Notes one such official, "We're here to develop teachers, not to screen people out. For the most part, everyone who enters the program is going to complete it, unless they decide that teaching's not for them."

The Costs of Certification

Especially for anyone who didn't complete a teacher-training program as an undergraduate, the costs of certification can be significant. It is not unusual for postgraduate teacher training programs to require a full-time commitment of 16 or even 24 months or a part-time commitment that can stretch to three years or more. The cost of training and the loss of salary due to time spent out of the workforce can easily reduce a teacher's real compensation during her first five years by 25 percent or more.

These barriers make other professions relatively more attractive, so that potentially talented teachers who are unsure about their interest are less likely to try teaching. Whereas candidates can readily try journalism or consulting or marketing for a year, they must make an extensive commitment before they can try public school teaching. The result is that many who might make fine teachers never enter the profession. There is disturbing evidence that certification may especially dissuade accomplished minority candidates—who have a number of attractive career options and who are often less well situated to absorb the costs of teacher preparation—from entering teaching.

This would pose no real problem if we were blessed with a surplus of good teachers. In such a case, we might scoff "good riddance" to those dissuaded from teaching. However, we have a desperate need for competent teachers. Moreover, rather than a lack of commitment to teaching, a reluctance to pursue certification may indicate that individuals have attractive alternatives. It is the most talented and hardest working individuals who have the most career options and who sacrifice the most by entering a profession where compensation is unlinked to performance and where opportunities for advancement are few. They may wish to teach but be unwilling to forgo work for a year, sit through poorly regarded courses, or jump through procedural hurdles. It is candidates with fewer attractive options who will find the tedious but intellectually undemanding requirements of certification less problematic. In fact, by suppressing the supply of teachers, certification provides teachers with enhanced job security. Coupled with a compensation scale that rewards seniority rather than performance, certification may well make the profession more attractive to graduates seeking a less demanding line of work. In this way, certification can actually harm the public's perception of teaching as a profession—the very opposite of what certification proponents wish to do.

Creative Destruction

In a world without certification as we know it, districts and schools would have more flexibility to ensure that their new teachers are prepared, inducted, and supervised in a manner appropriate to the challenges at hand. Because aspiring teachers would no longer have to attend formal teacher-preparation programs in order to teach, they would be free to make professional decisions about training in the same manner as business school or journalism school students. Weaker teacher-preparation programs would likely fall by the wayside. The fact that schools of education could no longer rely on a captive body of aspiring teachers would expose them to the cleansing winds of competition. Schools would have to contribute value—by providing teacher training, services, or research that created demand and attracted support—or face significant cutbacks. Teacher-preparation programs would find it in their own self-interest to ensure that their graduates were knowledgeable and skilled, as this would help graduates to win desirable jobs amid increased competition,

Professional educators themselves have been unable to explain in any concrete sense what teachers need to know and be able to do.

The Future Starts Today

There is disturbing evidence that certification may especially dissuade accomplished minority candidates—who have a number of attractive career options—from entering teaching.

making preparatory institutions more attractive.

Under a competitive certification system, little is likely to change in many of our high-performing suburban districts, where officials are inundated with applicants and are unlikely to tamper with a formula that is "working." In such districts, except in rare cases, we would expect that administrators would continue to cherry-pick from the nation's top teacher-education graduates. It is in the less desirable and more troubled systems, the nation's urban and rural school districts, that administrators currently have tremendous difficulty finding sufficient numbers of certified teachers. This is doubly true in the areas of math and science education. It is in these districts and subjects, where critics have fretted about the numbers of long-term substitutes and "burned out" veterans, where the wave of new teachers will most likely be recruited and welcomed.

Aspiring teachers would presumably continue to attend those teacher-training programs thought to enhance their employability.

While many of the resultant applicants will no doubt be deemed unprepared or unsuited for the jobs they pursue, there are few urban or rural principals who would not welcome the chance to pick and choose from their ranks.

Critics may fear that the elimination of licensure requirements will mean the end of teacher preparation and professional development. Such concern is unfounded. First, allowing uncertified individuals to become teachers does not mean that they must be viewed as "completed" professionals. Such a mindset is one of the vestiges of our current system, which is erected on a premise that all teachers are certified and therefore competent. Here, a better model might be medicine or law, where entering professionals begin their career with a trial period (serving as a hospital resident or as a junior partner in a law firm, for instance) during which their full panoply of skills is developed and monitored. Beginning teachers might serve on a probationary basis, receiving substantial monitoring and counseling. However, legal and contractual language ought to make it much simpler to terminate ineffective teachers or to mandate that they engage in support activities designed to improve their performance.

Second, moving to competitive certification does not mean doing away with professional teacher-education programs. Many applicants attend journalism school or business school, even though such training is not officially required, because it may make graduates more effective and can help them find better employment more readily. Likewise, aspiring teachers would presumably continue to attend those teacher-training programs thought to enhance their employability. This change would introduce some much-needed market pressure in this area, as schools would be forced to compete for students based on the usefulness of their course offerings.

Giving districts more leeway to hire promising candidates does not mean they will always make good decisions. Some ineffective teachers will inevitably continue to be hired. However, if entry to the profession is eased, it is appropriate that exit be eased as well. If administrators are to have more leeway to make hiring decisions, they also must be given more leeway to fire—and they must be held accountable for both sets of decisions.

At the end of the day, the individuals best equipped to assess the qualifications of prospective teachers are the principals who will be responsible for them. These same principals ought to have the strongest incentive to see that teachers are effective. If we believe that the administrators charged with managing and supervising schools either are unequipped to evaluate prospective teachers or are unwilling to do so, teacher certification will not suffice to protect our children from such profound systemic dysfunction. If we trust administrators, then certification is unnecessary and entails significant costs. If we don't trust administrators, let us address that issue directly and not rely on the hollow promise of flimsy parchment barriers. Regardless, it is past time to fully acknowledge the nuanced, multifaceted, and professional nature of teaching and to move beyond a system that restricts professional entry with procedural barriers characterized by ambiguity that are the result of an inability to clearly define the skills, knowledge, or training essential to good teaching.

–Frederick M. Hess is an assistant professor of education and government at the University of Virginia and directs the Virginia Center for Educational Policy Studies.

The FEDS STEP in

Will the new federal accountability standards spur school improvement?

FROM HIS FIRST DAYS IN OFFICE, President Bush made education reform one of his chief priorities. Congress responded with a reauthorization of the Elementary and Secondary Education Act (ESEA) that requires states to ensure that all students reach a certain level of proficiency within the next 12 years. Schools that fail to meet their achievement targets will face sanctions as harsh as being "reconstituted"—their staff being fired and forced to reapply for their jobs. The new legislation has enjoyed nearly universal bipartisan support, but the future remains uncertain. Are ESEA's goals realistic, or do they fundamentally misunderstand the nature of schools and learning? Do the states have the resources and the political will necessary to meet these more stringent requirements? And will school leaders take up the challenge or throw their hands up in frustration? No one has concrete answers, but one thing is sure: accountability is now the coin of the realm.

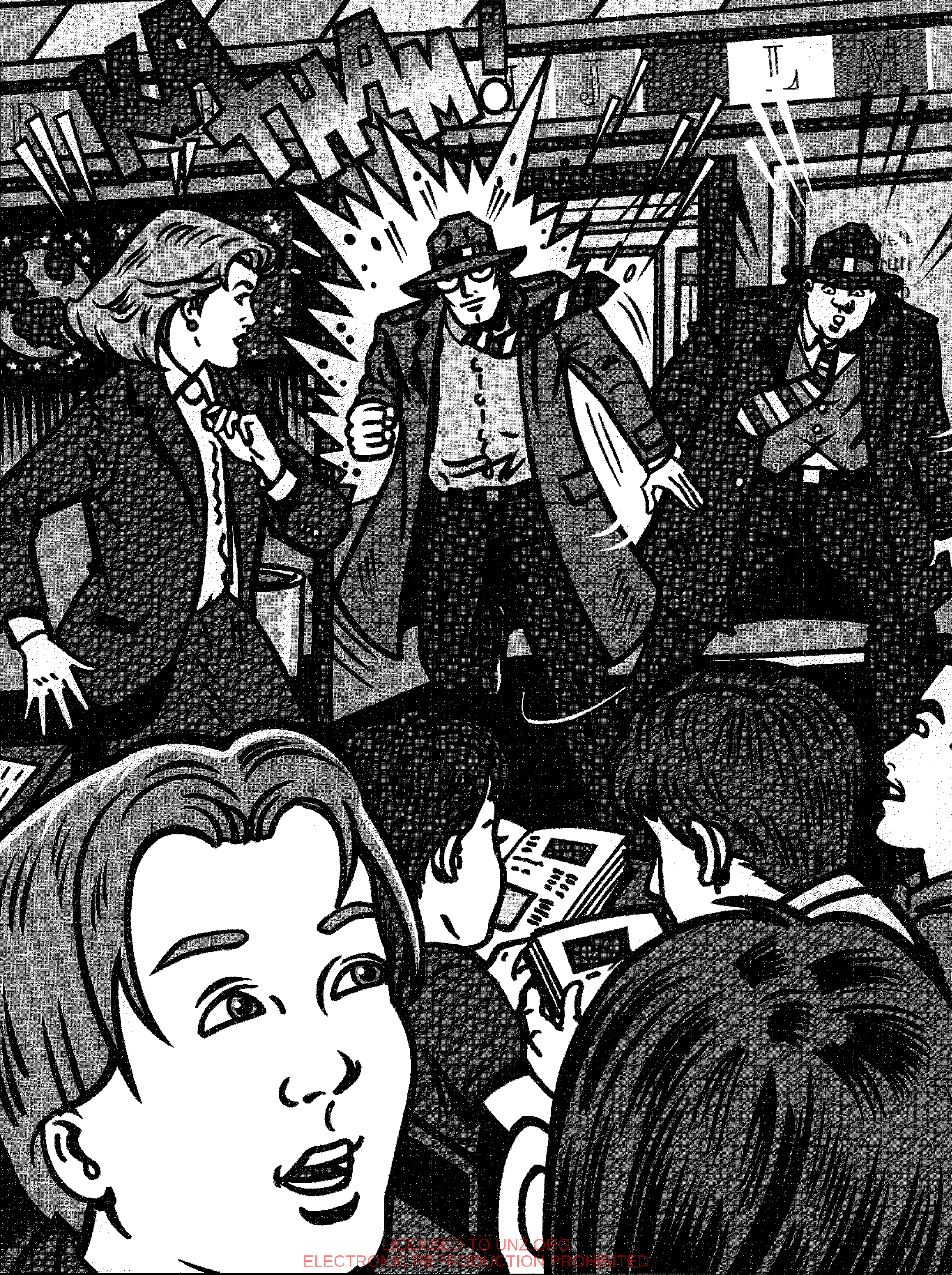
ILLUSTRATIONS BY ROBERT DALE

Richard F. Elmore doubts whether failing schools have any clue how to turn themselves around

PAGE 30

Andrew Rotherham wonders what the critics of accountability are trying to protect

PAGE 36


KA-THAM!
J
L
M

UNWARRANTED INTRUSION

Isn't it ironic that Republicans sponsored the single largest—and potentially the single most damaging—expansion of federal power over the nation's education system in history?

by RICHARD F. ELMORE

Inside the Washington, D.C., beltway, the reauthorization of the Elementary and Secondary Education Act (ESEA) is seen as either a sea change in federal education policy or a half-measure designed to demonstrate the political leadership's willingness to "do something" on education. On one side are supporters of the legislation who point to its substantial tightening of school accountability; its granting of more flexibility to states and school districts in the use of federal funds; and its commitment to applying sanctions to and providing aid for failing schools. On the other side are those who argue that the bill doesn't go far enough. Supporters of school choice believe that opportunities for real reform were lost when conservatives and New Democrats failed to persuade their colleagues and the president's advisors to include vouchers as part of the reform package.

In other words, there is no genuine opposition in Washington to accountability rules that simply fail to understand the institutional realities of accountability in states, districts, and schools. And the law's provisions are considerably at odds with the technical realities of test-based accountability. Never, I think, in the history of federal education policy has the disconnect between policy and practice been so evident, and possibly never so dangerous. What's particularly strange and ironic is that conservative Republicans control the White House and the House of Representatives, and they sponsored the single largest—and the single most damaging—expansion of federal power over the nation's education system in history.

The federal government is mandating a single test-based accountability system for all states—a system currently operating in fewer than half the states. The federal government is requiring annual testing at every grade level, and requiring states to disaggregate their test scores by racial and socioeconomic backgrounds—a system currently operating in only a handful of states and one that is fraught with technical difficulties. The federal government is mandating a single definition of adequate yearly progress, the amount by which schools must increase their test scores in order to avoid some sort of sanction—an issue that in the past has been decided jointly by states and the federal government. And the federal government has set a single target date by which all students must exceed a state-defined proficiency level—an issue that in the past has been left almost entirely to states and localities.

The accountability movement meant that politicians could take credit for improving schools without committing themselves to serious increases in funding.

Thus the federal government is now accelerating the worst trend of the current accountability movement: that performance-based accountability has come to mean testing, and testing alone. It doesn't have to. In fact, in the early stages of the current accountability movement, reformers had an expansive view of performance that included, in addition to tests, portfolios of students' work, teachers' evaluations of their students, student-initiated projects, and formal exhibitions of students' work. The comparative appeal of standardized tests is easy to see: they are relatively inexpensive to administer; can be mandated relatively easily; can be rapidly implemented; and deliver clear, visible results. However, relying only on standardized tests simply dodges the complicated questions of what tests actually measure and of how schools and students react when tests are the sole yardstick of performance.

If this shift in federal policy were based on the accumulated wisdom gained from experiences with accountability in states, districts, and schools, or if it were based on clear design principles that had some basis in practice, it might be worth the risk. In fact, however, this shift is based on little more than policy talk among people who know hardly anything about the institutional realities of accountability and even less about the problems of improving instruction in schools.

The Implementation Problem

The idea of performance-based accountability was introduced in the mid-1980s by the National Governors Association, led by Bill Clinton, then the governor of Arkansas. It took the form of what was then called the "horse trade"—states would grant schools and districts more flexibility in making decisions about what and how to teach in return for more accountability for academic performance. This idea became the central theory of today's accountability reforms. It was an appealing idea in principle: governors and state legislators could take credit for improving schools without committing themselves to serious increases in funding.

The movement got a major boost in 1994, when Title I—the flagship federal compensatory education program—was amended to require states to create performance-based accountability systems for schools. The vision behind the 1994 amendments was that Title I would complement and accelerate the trend that began at the state level; they required states to develop academic standards, assessments based on the standards, and progress goals for schools and school districts, all within ambitious timetables. The merger of state and federal accountability policies—or alignment, as it was called—was supposed to occur by the year 2000. By the end of the decade, it was difficult to find more than one or two states lacking some form of statewide testing program and public release of the results. However, in all but a few states the basic architecture of performance-based accountability systems remained relatively crude and underdeveloped. In those few states where the idea had been developed most extensively—Texas and Kentucky, for example—the systems worked well enough, according to the testimonials of their sponsors, to legitimate the idea that they were successful in general. Even in these states, however, there were legitimate criticisms of the accountability system's actual effect on academic performance and drop-out rates.

Nevertheless, by the late 1990s, it was abundantly clear that the states had fallen well short of what the crafters of the 1994 Title I amendments had envisioned. It was also clear that the federal government possessed very little leverage with which to force them along. States varied vastly in their administrative capacities to implement performance-based accountability systems. Many did not have testing systems in place that met the federal requirements. Most did not have the capacity to administer and monitor testing programs of the scale required to meet the federal goals. More important, creating accountability systems at the state level is essentially a political act, not an administrative one, and the federal government's harmless knuckle-rapping was hardly going to overcome the intransigence of a state legislature or governor. Indeed, the ability of the U.S. Department of Education to monitor and enforce compliance with the 1994 law was limited; budget cuts whittled away at the department's Title I staff just as their responsibilities were increasing. As is always the case, the department's senior political appointees were reluctant to make life too difficult for governors and chief state school officers, who are among their key political constituencies. The rub: by 2000, the target date for full compliance, fewer than half the states had met the requirements. In this environment, it came as no surprise to learn that by the year 2000 many schools with Title I–eligible students were simply unaware of the program's major policy shift in 1994.

This situation should have signaled to the Bush administration and Congress that there were complex issues of institutional capacity at the state and local level that could not be brushed aside by simply tightening the existing law's requirements. If more than half the states were unable or unwilling to comply with the requirements of the previous, less stringent, more forgiving law, why would one expect *all* the states to comply with a much more stringent and exacting law? Part of the problem is political. Even though virtually all the states have joined the accountability bandwagon, for many states it is largely a symbolic act. The basic designs of the systems are still primitive; the authority of state education officials to oversee school districts is still limited in many cases; and the political consequences of imposing large-scale, statewide testing in states with strong traditions of local control are risky. Consequently, support for accountability among state legislators and governors is often highly volatile.

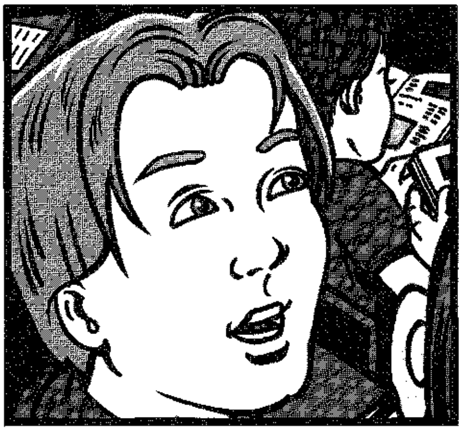
Another part of the problem is administrative. Mounting a statewide testing system is a task beyond the capacity of most state departments of education. Those that have embarked on large-scale testing are stretched to their limits just managing test-development work or monitoring testing contractors. Many states are not doing this work particularly well; others still don't know what the work entails. Still another part of the problem is technical. Standardized tests inevitably become highly politicized and, in the course of the debate, the limits of testing are subjected to public scrutiny. Many policymakers enter the accountability debate not knowing much about testing, and they often discover, much to their chagrin, that what they don't know can hurt them. Many legislators, for example, are surprised to hear that off-the-shelf standardized tests may not validly measure the content specified in state-mandated standards and that norm-referenced tests (tests that deliberately create a normal distribution around a mean) may not be effective in measuring changes in performance.

The Capacity Gap

The working theory behind test-based accountability is seemingly—perhaps fatally—simple. Students take tests that measure their academic performance in various subject areas. The results trigger certain consequences for students and schools—rewards, in the case of high performance, and sanctions for poor performance. Having stakes attached to test scores is supposed to create incentives for students and teachers to work harder and for school and district administrators to do a better job of monitoring their performance. If students, teachers, or schools are chronically low performing, presumably something more must be done—students must be denied diplomas or held back a grade; teachers or principals must be sanctioned or dismissed; and failing schools must be fixed or simply closed. The threat of such measures is supposed to be enough to motivate students and schools to ever-higher levels of performance.

This may have the ring of truth, but it is in fact a naïve, highly schematic, and oversimplified view of what it takes to improve student learning. The work that my colleagues and I have done on accountability suggests that internal accountability precedes external accountability. That is, school personnel must share a coherent, explicit set of norms and expectations about what a good school looks like before they can use signals from the outside to improve student learning. Giving test results to an incoherent, atomized, badly run school doesn't automatically make it a better school. The ability of a school to make improvements has to do with the beliefs, norms, expectations, and practices that people in the organization share, not with the kind of information they receive about their performance. Low-performing schools aren't coherent enough to respond to external demands for accountability.

The work of turning a school around entails improving the knowledge and skills of teachers—changing their knowledge of content and how to teach it—and helping them to understand where their students are in their academic development. Low-performing schools, and the people who work in them, don't know what to do. If they did, they would be doing it already. You can't improve a school's performance, or the performance of any teacher or student in it, without increasing the investment in teachers' knowledge, pedagogical skills, and understanding of students. This work can be influenced by an external accountability system, but it cannot be done by that system. Test scores don't tell us much of anything about these important domains; they provide a composite, undifferentiated signal about students' responses to a problem.



Test-based accountability without substantial investments in capacity—internal accountability and instructional improvement in schools—is unlikely to elicit better performance from low-performing students and schools. Furthermore, the increased pressure of test-based accountability, without substantial investments in capacity, is likely to aggravate the existing inequalities between low-performing and high-performing schools and students. Most high-performing schools simply reflect the social capital of their students; they are primarily schools with students of high socioeconomic status. Most low-performing schools also reflect the composition of their student populations. Performance-based accountability systems reward schools that work against the association between performance and socioeconomic status. However, most high-performing schools elicit higher performance by relying on the social capital of their students and families rather than on the internal capacity of the schools themselves. Most low-performing schools cannot rely on the social capital of students and families and instead must rely on their organizational capacity. Hence, with little or no investment in capacity, low-performing schools get worse relative to high-performing schools.

Some changes in the new law provide relatively unrestricted money that states can use to enhance capacity in schools if they choose to. However, neither state nor federal policy is currently addressing the capacity issue with anything like the intensity applied to the test-based accountability issue. So an enormous distortion is occurring in the relationship between accountability and capacity, a distortion that is being amplified rather than dampened by federal policy.

Abusing Tests

During the cold war, just about anyone who raised questions about the distribution of wealth in America was branded a Communist, thus chilling debates over social justice. Debate in the realm of education reform is being similarly chilled. Critics who suggest that there might be problems with the ways in which tests are being used for accountability purposes have been essentially marginalized. They're smeared with accusations of being against accountability of any kind and of being apologists for a broken system. The idea that the performance of students and schools can be accurately and reliably measured by test scores is an article of faith in test-based accountability systems. Consequently, tests are being misused and abused in ways that will eventually undermine the credibility of performance-based accountability systems.

Probably the most serious problem lies in the use of test scores to make decisions about students' academic progress—decisions about whether they can advance to the next grade or graduate from high school. The American Psychological Association's guidelines for test use, as well as the consensus of professional judgment in the field of educational testing and measurement, specifically prohibit basing any consequential judgment about an individual student on a single test score. The primary reason for this principle is technical, not ethical. Test scores have a significant margin of error associated with them. That margin of error increases as the number of cases decreases; individual scores are typically much less reliable than aggregates of many individual scores. The best that can be said about an individual test score is that it falls within a range that is described by its coefficient of reliability. Unless the range is extremely small, which it isn't for any standardized test, the likelihood of error is high.

The solution to this problem is to use multiple measures of a student's performance when making consequential decisions. But this solution is more expensive because it introduces a new level of technical complexity into the system. For instance, say that high-school graduation was based on a composite of grades, test scores, and portfolios of students' work. Developing such a composite would not only be a challenging technical feat; it would also introduce a certain amount of human judgment into the system. Policymakers tend to distrust the professionals who make such judgments.

A similar problem arises at the school level. Under Title I, schools are expected to meet their annual yearly progress goals. This involves calculating a school's annual gain in test scores from one year to the next. Title I also requires disaggregating these scores by students' ethnic and economic backgrounds. But research has shown (see Thomas Kane, Douglas Staiger, and Jeffrey Geppert, "Randomly Accountable," in this issue) that these measures are highly unreliable for schools the size of a typical elementary school, and they are particularly unreliable for even smaller groups of students. Schools are often misclassified as low or high performing purely because of random variation in their test scores, unrelated to any educational factor.

The standards and accountability movement is in danger of being transformed into the testing and accountability movement. States without the human and financial resources to select, administer, and monitor tests are now being forced to begin testing at all grade levels. This is the surest way to guarantee that the test will become the content. Instead of creating academic standards that drive the design of a standards-based assessment, low-capacity states will simply select a test based on its expense and ease of administration. Thus the criticism that many state assessments are invalid because they fail to test the curriculum that is being taught. The more the issue of validity is submerged in the political debate on accountability, the more likely that charges of "teaching to the test" will be essentially accurate. A test with no external anchor in standards or expectations about student learning becomes a curriculum in itself, which trivializes the whole idea of performance-based accountability.

> The standards and accountability movement is in danger of being transformed into the testing and accountability movement.

Prognosis

The idea of performance-based accountability plays to the greatest weaknesses of the American education system. After World War II, most industrialized countries nationalized their education systems, but not America. Just the idea that students in Louisiana should be held to the same academic and performance standards as those in New York was enough to inspire heated political debate for decades. One consequence of leaving decisions about content and performance to states and localities for so long is that they never developed the institutional capacity to monitor the improvement of teaching and learning in schools, to support the development of new knowledge and skill in teachers and administrators, and to develop measures of performance that are useful to educators and the public.

The difficult, uneven, and protracted slog toward clearer expectations and supports for learning has barely begun in most states and localities. The history of federal involvement in that long endeavor is at best mixed and at worst a failure. The current law repeats all of the strategic errors of the previous ESEA reauthorization, only this time at a higher level of federal intervention. The prognosis is not good. The best we can hope for is that the capacity problems of states and localities will become more visible as a political issue at the state and federal levels, triggering responses that will help schools overcome the obstacles they face in improving the quality and intensity of teaching and learning. Likewise, it is to be hoped that the technical failures of testing will gain higher visibility and trigger a response that focuses more on the assessment of student learning and less on the administration of tests. The worst that can happen is that test-based accountability will widen the gap between schools serving the well-off and the poor, thereby confirming, at least in the public's mind, that expecting high levels of learning from all children is unrealistic.

As with many policy innovations, performance-based accountability in education is mutating into a caricature of itself. Never has this been clearer than in the reauthorization of Title I. The idea of giving schools and school districts greater flexibility in return for greater accountability for student performance—the original principle behind the "horse trade" of the 1980s—makes a great deal of sense. What we have discovered, however, is that accountability for performance requires substantial investments in organizational capacity: state departments of education need the capacity to select, implement, and monitor sound measures of performance; schools need support in developing internal coherence and instructional capacity; schools and districts need help in creating reasonable, diverse ways of assessing student learning; and teachers need support in acquiring the knowledge and skill required to reach larger numbers of students with more demanding content. As performance-based accountability becomes test-based accountability, these critical issues recede, and a sensible policy becomes a nightmare.

–Richard F. Elmore is a professor of educational leadership at the Harvard Graduate School of Education and a senior research fellow at the Consortium for Policy Research in Education.

A NEW PARTNERSHIP

The move toward federally imposed accountability standards is necessary to ensure that federal funds are enhancing educational opportunity, especially for poor and minority students. It will all be for naught, however, if Congress doesn't guarantee that states will receive the resources necessary to overhaul failing schools

by ANDREW ROTHERHAM

The issue of whether the federal government should outline and enforce an accountability system for states, school districts, and schools was essentially settled the day that George W. Bush took office as president. Bush had made "accountability" a cornerstone of his education platform, using his stated goal of ensuring equity for poor and minority children as a way of bolstering his credentials as a moderate. New Democrats, led by Democratic senators Joseph Lieberman and Evan Bayh, were also committed to the idea of accountability. They had made a results-based approach to federal education programs a major component of their "Three R's" proposal—on which much of the Bush plan and the final ESEA legislation were based.

The legislation would build on the accountability measures first introduced during the 1994 reauthorization. That legislation required the states to develop academic standards and tests linked to the standards, but its accountability language was too vague and porous. For example, under the 1994 legislation, states were required to define "adequate yearly progress" in a way that resulted in "continuous and substantial yearly improvement" by schools and school districts toward the goal of getting all students to the proficient level. With states defining the annual yearly progress standard and with no concrete timeline in place, practices varied widely from state to state. This led to great differences in results at the state level. Michigan and Arkansas identified 76 percent and 64 percent, respectively, of their Title I schools as low performing, while Connecticut and Maryland identified only 6 percent each. Regardless, the low-performing label is sadly almost meaningless: 41 percent of principals in these schools reported not even being aware of the designation. And for schools that were identified as low performing there was little in the way of sustained assistance. According to a Department of Education analysis, only 40 percent of schools identified as needing improvement received assistance from the state or their school district. This dismal statistic climbs to only 50 percent for schools that have been identified as low performing for three or more years.

In this system, what passed for accountability was the ability to provide detailed reports of planned and actual spending of federal funds—in other words, a system of accounting, not of accountability. In addition, the federal government was quite lax in enforcing the accountability provisions that were in the law. As a result, as of this writing only 16 states had fully complied with the requirements of the 1994 ESEA reauthorization.

Critics have seized on the states' seeming inability to comply with the previous law as evidence of the folly of proceeding down the accountability route. In fact, critics warn, these federal efforts to demand results-based accountability are at best futile and at worst drive all sorts of perverse and unintended consequences, jeopardizing recent accomplishments at the state and local level. These arguments are the dullest yet most common arrows in the quivers of those fighting change. And they fall apart under close scrutiny. The critics' alternative to the accountability plan is to keep the federal dollars flowing regardless of the results. They have little to offer beyond tired bromides about needing more money for capacity building, innovative partnerships, and a host of other buzzwords that make no difference in the lives of children who attend failing schools.

Critics have little to offer beyond tired bromides about needing more money for "capacity building" and a host of other buzzwords that make no difference in the lives of children who attend failing schools.

Why Federal Accountability?

The states' haphazard results in complying with the 1994 requirements and improving low-performing schools overall are precisely why the law's accountability provisions ought to be strengthened and clarified. In a host of policy areas inside and outside of education, history shows that clear federal prescriptions accompanied by real consequences bring results. That's why, for example, you can't buy an alcoholic drink almost anywhere in the country if you're under 21. It's why our cars and airplanes are increasingly safe. And it's why the vestiges of discrimination are being eradicated from our schools and society, through laws like the Civil Rights Act and the Americans with Disabilities Act. Federal policymakers did not wait for states to address these issues on their own. Nor did they lament the states' incapacity to do so. Rather, they mandated clear standards and demanded results.

Increasing the rigor and specificity of the accountability provisions in ESEA doesn't mean imposing the same system on every state, or a "one size fits all" approach in the political jargon. It means establishing clear criteria for improvement, specific indicators of success, and common goals. Such a system need not be incompatible with a variety of approaches at the state level, but it is at odds with the notion that wildly divergent results for students of various races, ethnic groups, and incomes are somehow acceptable.

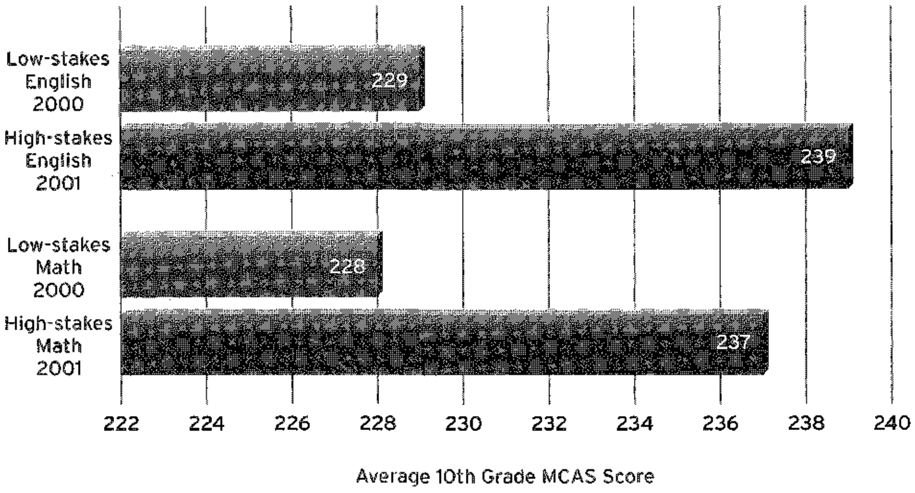
Obviously there is a world of difference between a standard that says all cars must contain passenger-side airbags by a certain date and one that defines what it means to be academically proficient. Complicated issues of how to measure success or failure vex the process of education policymaking. Still, complicated doesn't mean futile. It doesn't mean we should just throw up our hands in collective frustration, because in the end it is simply irresponsible to continue pouring resources into systems that we know are failing without establishing clear benchmarks for their improvement and consequences if they do not reach them. To do otherwise essentially makes Washington the enabler in a terribly dysfunctional relationship that victimizes poor and minority children. It is ironic that many of the same interest groups and individuals that so readily look to Washington to address various ills suddenly resist federal intrusion in this particular area. Are they satisfied with a situation where African-American and Hispanic 12th graders read and do math as well as white 8th graders? Where, according to an analysis by Jay Greene of the Manhattan Institute, only 56 percent of African-American students and 54 percent of Hispanics graduate from high school?

Perhaps it's because white students score higher on achievement tests and graduate at substantially higher rates that many of the loudest voices in this debate aren't troubled by asking for patience and time to get things exactly right before proceeding. These critics can represent a powerful bloc. Consider that in states like Massachusetts, Virginia, and New York, resistance to the accountability system has come predominantly from affluent white suburbs. Call it the Scarsdale Syndrome. In Massachusetts, writes Georgia Alexakis in the *Washington Monthly*, the paradox of these reform efforts is, "The schools most likely to do poorly on the MCAS [the state test in Massachusetts] have also been most likely to embrace it, while those districts whose scores are already quite high are fighting hardest to get rid of it." The relative political strengths in such a fight are sadly obvious; this is one more reason why accountability can't be left solely to the states.

Listening to the critics' complaints, one is left wondering, What are the wondrous accomplishments that more rigorous accountability will place in jeopardy? It is clear that we have failed to successfully educate poor and minority students on a large-scale basis. It is also clear that despite their best intentions many teachers, principals, superintendents, and professors at schools of education do not know how to address these shortcomings on a meaningful scale. Federal requirements driving states to address these problems would be much more troubling if they were interfering with a variety of successful state and local approaches, but that is simply not the case. In fact, based on what we know, federal accountability provisions will complement the most successful state practices.

The Pressure of Exit Exams (Figure 1)

In 2001, Massachusetts students needed to pass the 10th grade MCAS in order to graduate; that year, average test scores increased dramatically.



SOURCE: Massachusetts Department of Education

States Show Results

Will there be unintended consequences from the new federal accountability provisions? Undoubtedly. Will they all be perverse? No one knows, and unintended consequences that prove positive are certainly not unheard of. We do, however, know something of the intended consequences policymakers hope for. There is evidence that accountability systems with concrete goals change the behavior of school systems, at a minimum by refocusing efforts on disadvantaged students. Consider the experiences of Massachusetts, Virginia, and Texas. In Massachusetts and Virginia, where students are tested in key grades and will soon need to pass exit exams to graduate from high school, Cassandras predicted all kinds of pernicious results. They haven't materialized. In fact, although much work remains, test scores are steadily rising in both states.

During the past four years, the share of Virginia students passing the Algebra I and Algebra II Standards of Learning (SOL) assessments has risen by 34 and 43 percent, respectively. The pass rates for African-American students have gone from 20 percent to 59 percent in Algebra I and 13 percent to 58 percent in Algebra II since 1998. Virginia still needs to address a substantial achievement gap, but its minority students' scores have clearly improved.

In Massachusetts, test-score performance improved, once graduation requirements were imposed (see Figure 1). In 2001, 75 percent of 10th graders passed the math portion of the state's MCAS test, and 82 percent passed the language-arts test. This is up from 55 percent and 66 percent, respectively, the previous year. The Massachusetts example is particularly encouraging in light of Achieve, Inc.'s recent finding that "The grade 10 tests are rigorous yet reasonable—and are, in fact, the most challenging of the exit-level tests Achieve has yet reviewed." Achieve also lauded Massachusetts for its work to align its standards, curriculum, and assessments, which has provided a model for other states.

In Texas, where the TAAS test is widely considered to be less rigorous than the SOLs or MCAS tests (although the state is revising the TAAS), minority students nonetheless have shown gains that are corroborated by the National Assessment of Educational Progress. There is considerable evidence that during the past decade in Texas the needs of minority students have received increased attention as a result of an accountability system that demands that a school show not only overall progress, but also progress among its most disadvantaged charges.

Accountability systems are no panacea, and there are certainly problems in Texas and elsewhere. But these results at least indicate that accountability systems can help to focus attention on poor and minority students whose needs have been ignored or neglected. These results also seem to prove the point that states, if left to their own devices, will take action. It's true that the accountability movement has been state-led, to a large extent. Yet most states have yet to meet the requirements of the 1994 law, and it's clear that some won't move forward in any aggressive way without federal action.

What Must Be Done

Obviously, designing an accountability system of this nature is complex. Any workable proposal must be clear to practitioners; fair in the sense of not holding educators accountable for things they can't control; technically sound; and supported and enforced. It also should not squelch promising approaches that the states are developing.

Perhaps the most contentious issue in this debate is the use of standardized tests to measure school performance. Much of the hostility to accountability is actually just hostility toward testing. While standardized tests are certainly not perfect (in fact, they're primitive from a technological point of view), they're still the best objective way to measure progress. They lay bare discrepancies in educational quality in a quantifiable way. Sure, too many state assessment systems are lacking in quality or rigor. Yet this is a case for improvement, not abandonment.

There are legitimate complaints about the ways in which states are using the results of standardized tests. As Thomas Kane and Douglas Staiger point out (see "Randomly Accountable," in this issue), test scores bounce up and down from year-to-year for a variety of reasons that are unrelated to actual school performance. Thus no system should rely solely on the snapshot of a single year's test scores in making decisions about incentives or consequences. Accountability systems also need to include safeguards against the statistical unreliability of small classes and demographic groups that may include only a few students at a particular school.

Because some states are experimenting with value-added approaches to measuring school progress, it's important that federal accountability standards allow for this type of innovation. And while it may be desirable to have a purely technocratic system that makes no allowances for political and human impulses, it is not feasible. Because schools are human institutions shaped by a variety of forces and influences that may or may not be within their control, some "give" is required to address exceptional circumstances that will inevitably arise. States should have discretion to undertake and prioritize interventions and consequences. But such discretion (or "safe harbor" provisions) need not equal the vague 1994 language or allow states to use measures that are divorced from academic results or are purely subjective and porous. During the debate over the federal "annual yearly progress" standard, many of the proposals that would have included other indicators as measures of a school's annual yearly progress were simply thinly disguised attempts to eradicate any rigor from the system. There were proposals to, among other things, hold schools accountable only for the progress of the lowest-performing students in the bottom quintile; not disaggregate data by race and ethnicity; require states to deal only with the lowest-performing schools; or ignore test results altogether as an accountability tool.

The new law appears to have addressed all of these issues in a workable manner. In the end, it may well turn out that the president's mandate that states annually test all children in grades 3 through 8 will prove to be much more burdensome and troubling for states than the new accountability provisions. It's also entirely possible that the annual yearly progress provisions will cause trouble, as more and more schools wind up on lists of the low performing and politicians take the heat. Nonetheless, there is reason for cautious optimism.

What Washington must avoid is simply demanding accountability and then walking away. The New Democratic Lieberman-Bayh approach on education was predicated not only on more accountability, but also on more investment, more flexibility for states and localities, and a strategic federal role aimed at helping states and localities solve these problems, serving almost as a consultant to states and localities. This argues for a more active but less programmatic federal role in education.

States are increasingly failing to reach their revenue targets as a result of the slowing economy. The reforms of the ESEA legislation, especially the testing requirements, will require an expenditure of state resources on issues that aren't tied directly to the day-to-day provision of education. In a tight fiscal climate, testing and accountability initiatives will be curtailed or put on hold before direct services. It's up to Washington to help see that states aren't forced to make this choice. It is also essential that funding for interventions in low-performing schools accompany the new requirements. President Clinton inaugurated an accountability fund as part of Title I as a way to focus resources specifically on this purpose. However, more money will not provide a solution without enforcing clear goals for results.

There must also be a greater emphasis on getting these new resources to underserved communities. Funds must be concentrated rather than spread as far and wide as possible for political advantage. Democratic senator Mary Landrieu, a cosponsor of the "Three R's" bill, worked tirelessly, and against considerable opposition from members of both political parties, to increase the targeting of federal education dollars to low-income communities and schools in an effort to better support their school reform efforts.

Unfortunately, this legislation does not do enough to define the federal role in terms of consolidating programs and increasing local flexibility to meet diverse circumstances. The law includes flexibility provisions and some streamlining that are improvements. However, if raising overall test-score performance and addressing the achievement gap are to be the main focus of federal policy, it is foolish to have a panoply of programs that direct state and local officials toward a host of other priorities, distracting them from their core mission.

What Washington must avoid is simply demanding accountability and then walking away. It is essential that funding for interventions in low-performing schools accompany the new accountability requirements.

Someone must enforce the new rules if they are to be workable. Washington will encounter resistance at both the federal and state levels. As Senator Bayh of Indiana aptly told the *Los Angeles Times* in the midst of the debate, "Everyone is for accountability until it actually gets put into place and applies to them." Washington has a dismal record of enforcing its dictates in education. Of course, considering some of those dictates, sometimes this is a blessing. But one lesson of the 1994 reforms is that without enforcement states will simply ignore or delay parts of federal education laws they don't like. Tightening and clarifying the accountability provisions and then failing to enforce them only means states will be ignoring a new set of requirements. That's not much of an improvement on the status quo.

It's worth remembering that an army of naysayers predicting adverse consequences, or at best futility, has accompanied every major federal policy shift in education. However, the positive impact of accountability systems, particularly for the poor and minority students who traditionally have been excluded from educational opportunities, outweighs the risks. Mistakes will be made, lessons will be learned, policies will be fine-tuned. But we shouldn't delay the good while waiting for the perfect.

–Andrew Rotherham is the director of education policy at the Progressive Policy Institute in Washington, D.C.

In California, alternative certification has become a crucial source of new teachers. It would be even more crucial if its cumbersome requirements didn't dissuade so many promising candidates

tortuous routes

by DAVID RUENZEL

THE URBAN SCHOOL DISTRICTS OF CALIFORNIA HAVE a well-publicized shortage of teachers. So they're eager to move well-qualified candidates into the classroom, right? Not always. Nontraditional candidates—namely recent college graduates and career changers who haven't attended a standard teacher-preparation program—often encounter serious roadblocks, even with the state's full endorsement of alternative certification programs that allow candidates to start teaching while working toward their certification requirements.

Take Eddie Wexler. Just a few credits and a dissertation short of receiving his doctorate in education policy at the University of California at Berkeley, he abandoned his graduate work in order to become a high-school history teacher in the Oakland Public Schools. He figured that his substantial coursework in education and his experience as an assistant teacher in a Montessori school, as a school psychologist intern at a low-income urban school, and as an instructor at Berkeley would make him a viable candidate.

In late February of 1998, Wexler arrived at the offices of Oakland Unified with his transcript, three letters of recommendation, and his fingerprinting clearance from the county. At the time, Oakland claimed to have such difficulty finding qualified candidates that 16 percent of its teachers were working on emergency permits and waivers. In California, these are teachers with, at best, a bachelor's degree and a passing score on the CBEST, the state's minimum competency test for teaching candidates. Yet the district was indifferent to Wexler's inquiries.

Wexler later passed the challenging Praxis exam in history,

ILLUSTRATION BY ALLAN DAVEY/MASTERFILE

"There is a terrible teacher shortage," says California high-school teacher Eddie Wexler, "but it's exacerbated by policies and paperwork that undermine the ability of good candidates to get a job in high-need schools."

which virtually all nontraditional candidates must pass in order to receive a preliminary credential. Fewer than half of all candidates pass on their first try. Before receiving his results, he called a dozen school districts in the San Francisco Bay area, many suffering from shortages in all areas. The only one to respond was in Contra Costa County, northeast of Berkeley. In April he began working there as a substitute teacher. District officials told him that he needed to enroll in a state-approved teacher-preparation program, which he would eventually have to complete, in order to have a chance of teaching full time in the fall on an emergency permit. Wexler protested, saying that many of the courses such programs required closely matched the very rigorous ones he had already taken at Berkeley. But the district, like all of the districts he contacted, said that it didn't matter; the course titles and numbers didn't match, and that was that.

Wexler finally enrolled in a 12-week education course at a private college—at a cost of $800—as a way of demonstrating his commitment. There he learned such things as the importance of taking your grade book with you while evacuating during a fire drill. Meanwhile, the district in which he was subbing simply told him to keep checking back. By August, he had not yet received a single phone call from an interested district.

Wexler's luck changed only when, on the advice of another teacher, he skirted the district bureaucracy and began to call principals directly. He also placed his calls the week before classes were to begin, the best time to seek a job as an emergency credentialed teacher. He finally hit pay dirt when he reached the principal of a high school in Richmond, a gritty city in Contra Costa County girded by oil refineries. She asked Wexler if he could be at the school in half an hour for an interview. "We met for thirty minutes, during which I spoke for about two minutes," Wexler recalls. "I was then hired on the spot on an emergency permit. I was so lucky to be at the right place at the right time."

Three years later, Wexler chairs the history department, coaches the debate team, serves as a mentor teacher, and is now, relatively speaking, a savvy veteran in a school where 38 out of 55 teachers left last year.

Wexler just received his preliminary credential, which required that he pass a test on the U.S. Constitution (required of all California public school teachers) and complete more coursework (taken at San Francisco State) in addition to the CBEST and subject-matter tests. The preliminary credential is valid for five years, during which he must work toward a "clear" credential. This requires that he complete a computer-education course, a special-education course, a course for CPR certification, and additional professional coursework.

Reflecting on his job search, Wexler expresses amusement at his naïveté. "I had given myself six months to find a job, which I thought was plenty of time. I mean, all I had been hearing about is this terrible teacher shortage. Well, there is a terrible teacher shortage, but it's exacerbated by policies and paperwork that undermine the ability of good candidates to get a job in high-need schools."

Wexler acknowledges that he could have applied to one of California's many university- or school district-based alternative certification programs, but he didn't think that the path to a credential would have been much easier. He still would have had to sit through hundreds of hours of coursework and workshops, most of which he felt were redundant in light of his background. The fact is that the routes of most alternatively certified teachers have been larded with almost as many superfluous, cumbersome, and costly requirements as traditionally certified teachers face. All for the privilege of earning a fraction of their former salaries, some of which they will then use to purchase the classroom supplies (basics like chalk, paper, and books) that districts often fail to provide.

Emily Feistritzer, president of the National Center for Education Information (NCEI) and an expert on alternative certification programs, says that Wexler's story is typical, both nationally and in California. "The teacher shortage is a paradox, really a paradox—something created by the education bureaucracy. We need to minimize the obstacles to entry and get these people into classrooms right away, working with mentor teachers."

Setting the Bar High

In California, as in most other states, alternative certification was created primarily to address teacher shortages. It began in California during the baby boom era. In 1955 Stanford and Claremont, with the help of a Ford Foundation grant, offered the first college-based internship programs. Candidates with a bachelor's degree went through, much as they do now, a summer training program and then started teaching full time in the fall. In 1967 these programs were sanctioned by state law, ensuring their expansion. In 1983, as California's student population soared, the legislature permitted districts to start their own alternative certification programs. This enabled hard-pressed districts such as Los Angeles and San Diego to move candidates quickly into their classrooms (See Figure 1 for the credential status of California's teachers).

Both university- and district-based alternative certification programs are now well established in the state. Nevertheless, critics of alternative certification programs continue to claim that such programs are quick and largely ineffective fixes for teacher shortages. They insist that teachers are best trained in traditional programs based in education schools. Nonetheless, alternative certification's success in recruiting and retaining talented, accomplished people like Wexler is undeniable. According to Feistritzer's organization, 40 states now provide alternative routes into teaching, up from only 8 in 1983. Nationwide, about 85 percent of alternatively certified teachers remain in teaching after five years. In California the rate is 84 percent—which is remarkable, considering that alternatively certified teachers work mainly in urban districts with notoriously high attrition rates.

Alternative certification also attracts a more diverse pool of candidates to teaching. Michael McKibben of the California Commission on Teacher Credentialing (CTC) reports that ethnic minorities compose 46 percent of those entering alternative programs; 27 percent are males who want to teach in elementary schools—three times the normal rate. Sixty percent have come to teaching from other careers. Altogether, alternative certification programs produced nearly a *third* (7,200, to be exact) of California's 24,000 new teachers last year. About 2,300 of these teachers went through one of the state's eight school-district intern programs, the largest by far of which is that of Los Angeles Unified. Most of the other teachers went through one of the state's 74 university-based internship programs.

The programs may last one to two years, but only one program, according to the CTC, is as short as the statutory minimum of one year. The CTC reports that all require at least 120 hours of preservice preparation in child development and methods before the candidate begins full-time teaching. Interns must also be mentored by a certified teacher for the duration of the internship, though it is hardly a secret that the mentoring is frequently less than thorough—needy urban schools, where interns usually end up, typically lack enough veteran teachers for an effective mentoring program.

During the school year, nearly half of the programs offer instruction one afternoon a week in a three-to-four-hour block; slightly more than one-fourth meet two afternoons a week. The Los Angeles School District, which prepares about 85 percent of the state's district interns, requires 120 clock hours of preservice, four-hour classes once a week after school (for one year), and an additional 64 hours of training during the second summer.

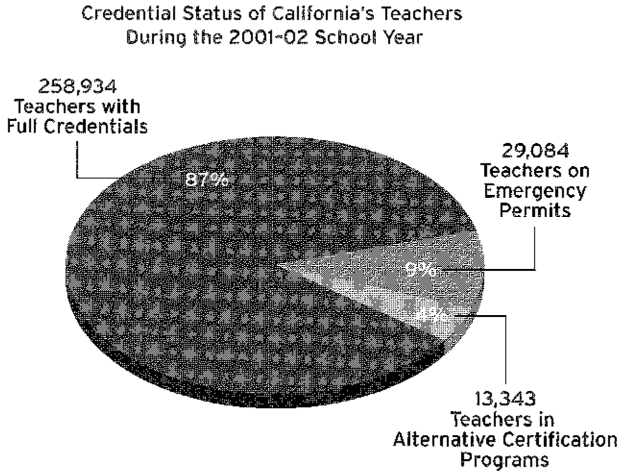
Within the broad framework outlined by the legislature, districts and universities have substantial leeway in structuring their programs. This makes generalizations regarding their character and quality difficult. Districts claim that their programs have the advantage of being extremely hands on and practical, while university program administrators often assert—though not for attribution—that district programs are shallow and insufficiently theoretical.

Alternatively certified teachers include engineers, biologists, lawyers, insurance representatives, and pediatricians. The quality of the applicants, says Anthony Lepire, chair of the department of secondary education at San Francisco State, is higher than it's ever been. "One of the things that's been uplifting for me after having been in the business for thirty years is the wonderful crop of people coming into teaching because they really want to be teachers. Most say, 'I have to do something that saturates my soul—I want to give back.'"

Alternative certification has clearly played a vital role in staffing California's schools with high-quality teachers and in easing the barriers to becoming a teacher. The sheer numbers speak for themselves. Nevertheless, becoming certified still requires enormous effort, even for candidates who have spent years teaching, but not in the public sector. Indeed, the impressive retention rates of alternatively certified teachers may say less about the programs themselves than about the fact that they have attracted candidates who so desperately want to become teachers that they will overcome any obstacle in their path. It is impossible to know how many good candidates have been *dissuaded* from becoming teachers because of the

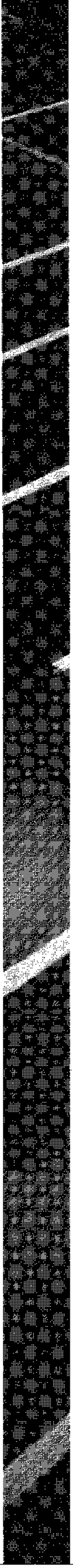
Drop in the Bucket (Figure 1)

The requirements of alternative certification programs in California are nearly as cumbersome as those of the traditional route to certification. As a result, nearly 10 percent of the state's teachers are still working on emergency permits.



SOURCE: Center for the Future of Teaching and Learning

In California, someone may have taught science for ten years, but if that experience came in a private school, it's as if it didn't even happen.



maze of requirements involved. Many of the alternatively certified teachers I spoke with told me that they have friends who wanted to become teachers but were deterred by the number of bureaucratic hoops they would have had to jump through.

California's alternative certification programs have the virtue of getting people into classrooms right away at a beginner's full salary, a clear improvement over the previous system. The problem is that they do not exempt teachers from any of the CTC's credentialing requirements. To get a clear credential, alternatively certified teachers, like traditionally prepared teachers, must pass a test on the U.S. Constitution, take a health-education course leading to CPR certification, and computer-education coursework. The intern's background makes no difference. Engineers, for instance, must take the computer coursework, which many describe as superficial at best.

Public school principals are still at a disadvantage vis-à-vis their peers in private schools: they don't have the ability to judge for themselves whether a degree from an education school or required coursework is important. In other words, they can't take a clearly bright candidate like Eddie Wexler and give him the experience and mentoring he needs to become a great teacher.

Those seeking alternative certification must pass subject-matter exams that are not required of traditionally prepared teachers, who have taken an approved sequence of courses in their majors. Non-traditional teachers have to do this before they can be accepted into internship programs. The failure rates on these exams, especially at the secondary level, are extremely high, even for those who have been successful in related careers. In 1996, 40 percent of whites passed the exam, 26 percent of Latinos, and 18 percent of African-Americans. Stories of teachers who have spent a year or two taking and retaking the tests are legion. While teachers certainly need to know their subjects, it seems that the bar is being set much higher for alternatively certified teachers than it is for those who attended education schools.

"I tell people that you should take the tests the first time to see what you're in for, because you're not going to pass them," said Lisa Rath, who teaches environmental science and biology at Birmingham High School in Los Angeles. A former private school teacher and corporate trainer who passed on her second try, Rath defends the test as a necessary tool for screening candidates.

Rich Hedman, an engineer turned high-school teacher in Sacramento, said it took him a year and a half to finally pass the exams. "I have degrees in both mechanical and aeronautical engineering, but they still wouldn't clear me to teach science," says Hedman. "Now that I've been teaching for five years I understand why—my background was all in applied science, and I needed to know more about biology and geology. Still, I'm not convinced that the tests accurately reflected my knowledge of science."

Some of the teachers I spoke with were resentful at having to pass exams they see as repetitive and irrelevant in some cases. Heather Nevis, for example, was a history major in college and a corporate lawyer in San Francisco, but she still had to take the history exams in order to teach high-school economics and psychology in the San Francisco Bay area. She has been teaching full time on an emergency permit for two years because she still must pass the multiple-choice section of the test. "To me, it's just more hurdles put in place by people who don't think teachers are smart enough. For me to have to pass a multiple-choice test with questions on the economic systems of ancient Egypt and Greece just doesn't make sense."

Perhaps would-be teachers should be expected to pass a difficult test of their subject-matter knowledge. If so, shouldn't *all* teachers be required to meet this rigorous standard? It would be quite illuminating to see what percentage of traditionally certified teachers could pass these tests. In any case, as McKibben notes, the difficulty of the subject-matter tests necessarily limits the expansion of alternative certification programs. "We have 37,000 teachers on emergency permits, partially on account of subject-matter concerns. When large numbers of teachers cannot pass the tests, schools are almost compelled to hire them on emergency permits."

Rigid Requirements

Passing these exams is only the beginning of an alternatively certified teacher's labors. As noted earlier, they must endure at least a year, and usually two, of intensive seminars, workshops, and courses—all while most are teaching for the first time in highly challenging urban schools. This is made worse by the fact that mid-career switchers, who compose the majority of interns, often lead busy family lives. "Internships aren't for everyone," McKibben said. "You have to be a mature individual to handle the pressure."

Lepire of San Francisco State, California's second largest intern program, said, "Our teachers can complete the program in two semesters or two years—whatever fits into their insanity best. Full time means 16 units per semester. In either case, you're barely surviving with coursework and seminars in the evening. You have no personal life."

Few would argue that teaching is simple work, that no professional coursework is necessary. Indeed, for some individuals—especially for those whose past experience has little connection with the skills required in teaching—the time commitments of the alternative certification programs may be perfectly reasonable. It's their inflexibility that makes them less of an alternative than they could be. A pediatrician may have spent decades treating children, but he or she will still have to take the full slate of child-development courses. Someone may have taught science for ten years, but if that experience came in a private school, it's as if it didn't even happen.

This inflexibility applies to salaries as well. The fact that new teachers entering an alternative certification program must start at the bottom of the salary scale, regardless of any related experience they may have had in other careers, undoubtedly limits the drawing power of such programs. Many veteran engineers interested in becoming math teachers—and math teachers are desperately needed—will likely blanch at the prospect of receiving the same pay as a 22-year-old just out of school. And the thousands of experienced private school teachers, a particularly rich source of expertise, are virtually shut out of public education altogether. Should they make the switch, their modest pay would become even more modest; they would also face the same credentialing requirements as those who have never taught a day in their lives.

The point is that alternative certification is, like so much in public education, bureaucratic and unnecessarily restrictive. There is no such thing as "holistic hiring," as is found in businesses, colleges, and private schools. The administrator ultimately responsible for hiring a teacher lacks the ability to say, "Your experience suggests that you need to fulfill this requirement, but these others can be waived." Alternative certification, like traditional certification, is still a matter of stipulations, credentials, and paperwork, to which no exceptions can be made.

Selective Quoting

An obstacle to both the expansion of alternative certification programs and the easing of their credentialing requirements is the widely held belief that there are few if any good substitutes for traditional teacher-education programs. No one has argued this more intensely than Linda Darling-Hammond, a professor of education at Stanford University and former executive director of the National Commission on Teaching & America's Future. While she is no longer as dismissive of alternative certification programs as she once was, she regards all but a very few—those with the most requirements—with suspicion.

Darling-Hammond's views of alternative certification became infamous with the 1994 publication of a long and vituperative article she wrote for *Phi Delta Kappan*. The article savaged Teach for America and its founder, Wendy Kopp, for placing new college graduates into urban schools with only a few weeks of hastily conceived training. Darling-Hammond attacked Teach for America for many things, but especially for its belief that good teachers are made primarily through experience. This belief is the cornerstone of all alternative programs.

As an example of why intense training is needed to become a teacher, Darling-Hammond cited the case of Jonathon Schorr, a Teach for America member who graduated from Yale. In a 1993 issue of the *Kappan*, Schorr had described how poorly prepared he was for the rigors of teaching in a Los Angeles school. However, Darling-Hammond quoted Schorr very selectively, using only those quotes that were critical of Teach for America. While he was teaching, Schorr also took a full load of education classes and concluded, "After two years of so-called teacher education, I can say that virtually nothing of what I know came from these classes.... Clearly, traditional teacher education does not hold the answers."

Surveys conducted by Emily Feistritzer suggest that Schorr's perspective is shared by a majority of teachers. Teachers consistently ranked experience as the most important factor to learning how to teach. They learn about teaching by teaching. Their second most valuable experience is working with other teachers. Way down on the list is education-school courses. This information, Feistritzer concluded, "makes a strong case for taking people with basic subject competency and getting them into classrooms and working with mentor teachers as quickly as possible."

The inflexibility of alternative certification applies to salaries as well. Veteran engineers interested in becoming math teachers will likely blanch at the prospect of receiving the same pay as a 22-year-old just out of school.

Most teachers I spoke with, both traditionally and alternatively certified, were decidedly unenthusiastic about their coursework experiences. Lisa Rath calls the education courses she took at California State–Northridge "a horrible waste of time, something out of the stone age—all theory with no applicability to the classroom." Only when she left the traditional teacher-preparation program and entered the Los Angeles district's

One teacher calls the education courses she took at California State–Northridge "a horrible waste of time, something out of the stone age—all theory with no applicability to the classroom."

intern program did she get practical experience, on which she was able to reflect with mentor teachers.

"The ed school courses I took at Sacramento State didn't at all prepare me for what I got myself into," says Michelle Maranta, an HIV researcher at the bio-tech firm Genentech who taught biology for one semester at a low-income school in South San Francisco before returning to the company. "I was overwhelmed with 220 students a day and the planning killed me. The best scenario would have been for me to get training while there, in the school, with maybe some team teaching."

"I don't submit that teaching is so easy that you shouldn't take any kind of class work," said Heather Nevis, the corporate lawyer turned social studies teacher who has done her teacher preparation at the private Chapman University. "But the hard part of teaching is classroom management and developing lesson plans that will reach as many students as possible, and you don't get that in class work. The real focus needs to be on learning while you're teaching in the classroom."

So why is there so much opposition to alternative certification? I asked this question of Michael Podgursky, an economist at the University of Missouri and a prominent critic of certification laws. I had expected him to point to economic incentives. Education schools, for instance, benefit financially when candidates must enroll in their programs. While he said these were a factor, he also suggested that there was a more important one. "Many people like Darling-Hammond sincerely believe that alternative certification will water down the profession," says Podgursky. "They think that teaching should be like medicine, with higher education—and only higher education—providing the training. But teaching isn't like medicine. While effective teaching requires a lot of knowledge and skill, it just isn't as rooted in science as medicine is." Podgursky is not optimistic about the future of alternative certification. "Demographics show us that student enrollment will level off by 2010, even in California. I fear that some states will shut down their alternative certification programs as the demand for teachers is greatly reduced."

Help Wanted

Under the current system, any effort to meet the teacher shortage by placing alternatively certified teachers in urban schools is bound to fall short. Los Angeles, despite its ambitious program, still had almost 6,000 teachers, or 19 percent, working on emergency permits and waivers in 1996. (While more current figures are not available, no one suggests they have changed much.) Far more Los Angeles teachers work on emergency permits than on district internships; in 1998–99, about 1,900 were working as district interns. In all of California, more than 34,000 teachers served on emergency permits in 1998–99, the last year for which complete data are available. This was about 12 percent of the state's total teaching staff of 276,000. About half of emergency credentialed teachers leave teaching within three years, usually to be replaced by more of the same (emergency permits must be renewed each year, up to a five-year maximum).

While not a complete solution, relaxing certification requirements for both traditionally and alternatively prepared teachers would certainly go a long way toward filling open positions with qualified candidates. The most promising candidates—those who have been successful in other careers—are least likely to embark on a tortuous route.

I asked Dennis Tierney, for many years the director of professional services at the California Commission on Teacher Credentialing and now a consultant, what I would have to do in order to teach in a California public school. I told him that I spent five years as an English department chair at a private school and more than ten years as a journalist covering education. He answered, "There are lots of programs designed for people like you. You do have to pass the CBEST, be fingerprinted by the Department of Justice and FBI, and demonstrate subject-matter competence, which you can do by taking two tests. Then, after a summer of coursework, you can teach as an intern while you complete requirements toward your credential."

Tierney spoke as if all these steps seemed perfectly sensible, but to me they seemed onerous, as I was sure they would to many others considering a mid-career change. Why would someone like me, with so many years of experience in education, want to fulfill what seemed like a series of redundant requirements? My years as an English teacher in a private school would count for nothing. I would start at the bottom of the salary scale unless a principal could somehow finagle me some extra dollars. I would also have before me a year or two of intensive coursework while working under the guidance of a certified teacher. For me, as for many others, alternative certification has yet to become a true alternative.

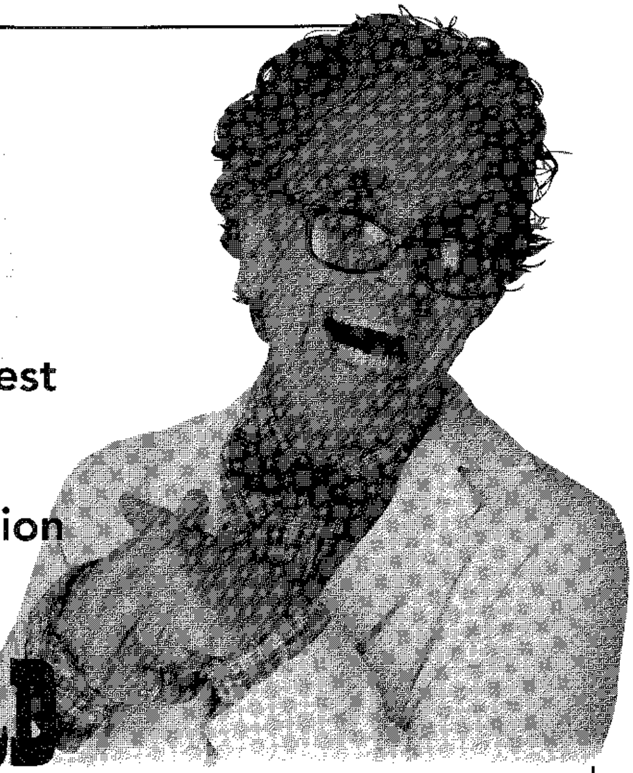
–David Ruenzel is a contributing editor and writer at Education Week *and* Teacher Magazine. *He also teaches writing and literature at a private school in the San Francisco Bay area.*

WANTED

For distinguished work in the quest for excellence in education

The Thomas B. Fordham Foundation is offering a

$25,000 REWARD



ALIASES: Governor, superintendent, education scholar, analyst, activist, policy maker, school founder, teacher, principal

DESCRIPTION: Distinguished education scholar, practitioner or policy leader who significantly advances the cause of education reform by exemplary work on behalf of higher academic standards, accountability, better teaching, and greater competition and choice

VIOLATION: Thinking outside the box about education reform

CAUTION: This person should be considered dangerous to traditional assumptions about education

If you have seen this person, please immediately contact the Thomas B. Fordham Foundation.

The Thomas B. Fordham Foundation announces two new prizes of $25,000 each to leaders in education reform. The prizes will be awarded annually.

THOMAS B.
FORDHAM
FOUNDATION

For details about these prizes and instructions for nominating a person for either award, please visit our web site at www.edexcellence.net.

The evidence shows that good teachers make a clear difference in student achievement. The problem is that we don't really know what makes A GOOD TEACHER

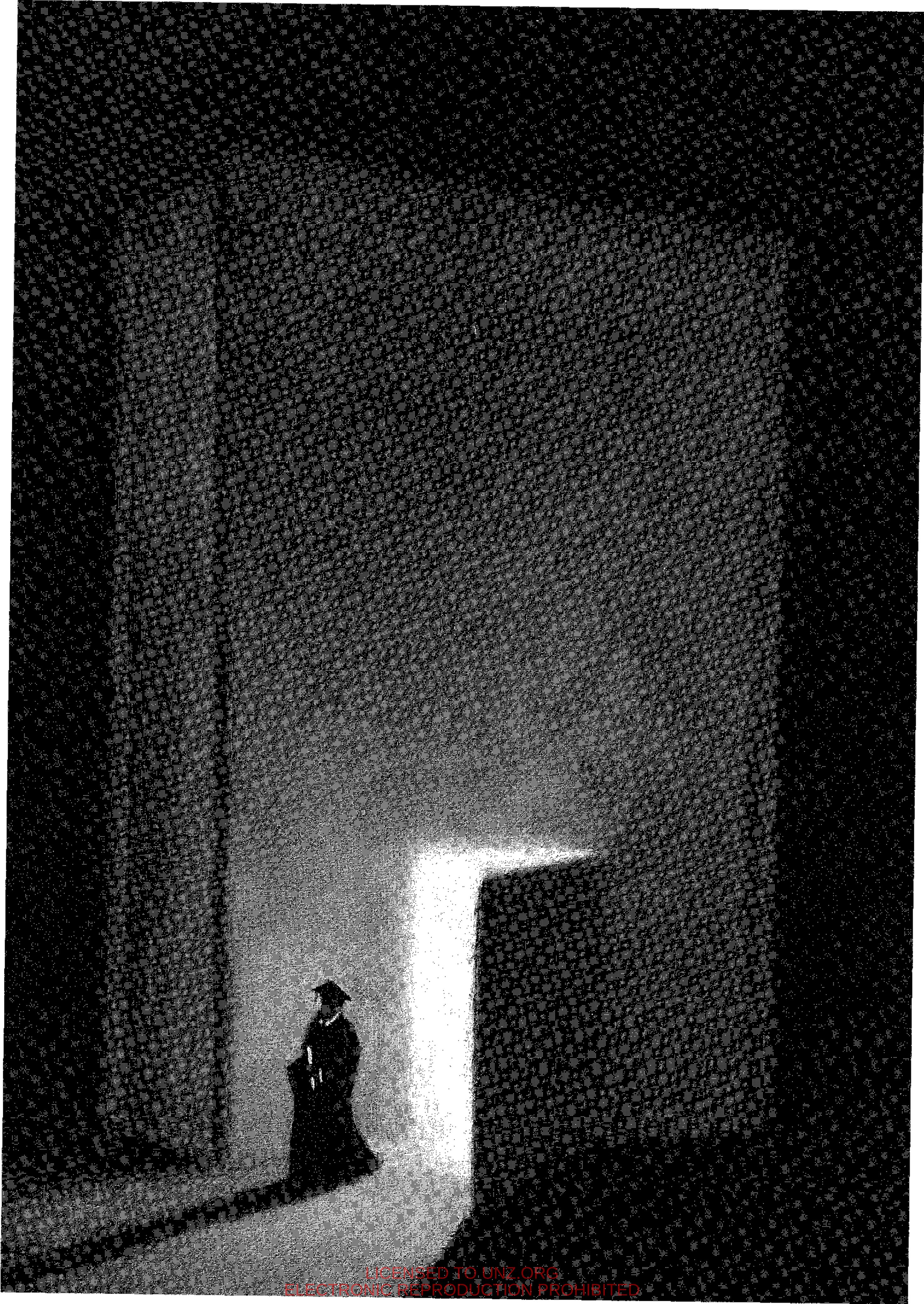
the Mystery of GOOD Teaching

by DAN GOLDHABER

WHO SHOULD BE RECRUITED TO fill the two to three million K–12 teaching positions projected to come open during the next decade? What kinds of knowledge and training should these new recruits have? These are the questions confronting policymakers as a generation of teachers retires at the same time that the so-called baby boom echo is making its way through the education system. Key to answering these questions is knowing how much influence teachers have over student achievement and what specific teacher attributes lead to higher student achievement. For instance, does holding a master's degree make one a better teacher? Do the best teachers hail from elite universities? Did they earn high GPAs in college? Did they major in the subject they are teaching? How much does experience matter? Do traditional, university-based teacher-preparation programs produce the best teachers, or are alternatively certified teachers just as good?

These questions are particularly relevant given that researchers have raised concerns about the overall quality of today's teaching workforce. As measured by standardized test scores (mainly the SAT and the ACT), students choosing to major in education tend to be drawn from the lower end of the ability distribution. In *Who Will Teach?*, Harvard University professor of education Richard Murnane and his colleagues write: "College graduates with high test scores are less likely to take [teaching] jobs, employed teachers with high test scores are less likely to stay, and former teachers with high test scores are less likely to return." On average, according to the

ILLUSTRATION BY ANDREW JUDD/MASTERFILE



findings of University of Massachusetts economist Dale Ballou, the higher the quality of an individual's undergraduate institution, the less likely a student is to choose a teaching career. Moreover, during the past 25 years the share of master's and doctoral degrees in education granted by top-tier public and private research universities has declined dramatically. And of students who graduated from college in 1993 and 1994, data from the Baccalaureate and Beyond survey show that those who entered the public school teaching profession averaged a 923 on the SATs; the average SAT of those entering other professions was about 80 points higher. The results are even more dramatic when one compares the SATs of teachers with those of people entering technical professions, such as engineering.

The demographic shifts of the next decade provide the opportunity to thoroughly remake the teaching profession. This is a significant opportunity, given that the evidence suggests that teacher quality is the most important school factor in explaining differences in student performance. The difficulty from a policy perspective is that the relationship between readily quantifiable attributes—such as a teacher's highest degree attained or level of experience—and student outcomes is tenuous at best. In other words, good teachers certainly make a difference, but it's unclear what makes for a good teacher.

Overall Impact

Research dating back to the 1966 release of *Equality of Educational Opportunity* (the "Coleman Report") shows that student performance is only weakly related to school quality. The report concluded that students' socioeconomic background was a far more influential factor. However, among the various influences that schools and policymakers can control, teacher quality was found to account for a larger portion of the variation in student test scores than all other characteristics of a school, excluding the composition of the student body (so-called peer effects).

The difficulty from a policy perspective is that the relationship between easily measured attributes—such as a teacher's highest degree attained or level of experience—and student outcomes is TENUOUS AT BEST.

Much of the research published since the Coleman Report has confirmed the finding that high-quality teachers raise student performance—indeed, it appears that the most important thing a school can do is to provide its students with good teachers. The Coleman Report's finding was based on the influence of a set of quantifiable teacher characteristics, such as years of experience, education levels, and performance on a vocabulary test. Since then, due in large part to the availability of new data sources that link and track teachers and students over a number of years, researchers have been able to estimate the *overall* contribution of teachers to student learning. This includes not only the effect of easily measurable attributes, such as experience and degrees obtained, but also the effect of harder to measure intangible attributes, such as a teacher's enthusiasm and skill in conveying knowledge.

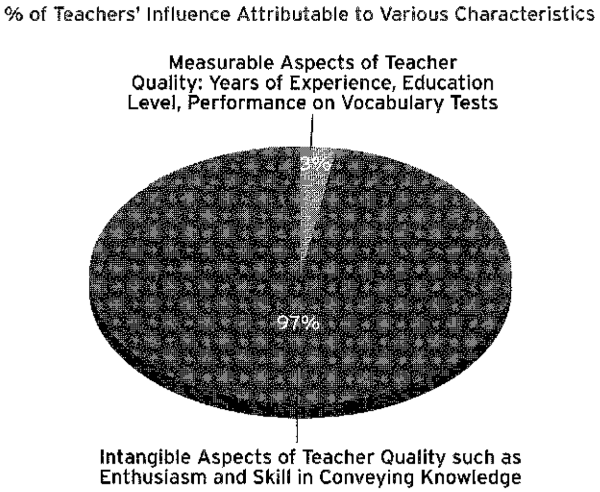
Tennessee is one of the few states with data systems in place that track teachers over time and link them to their students' achievement scores. Researchers who have studied the Tennessee data, mainly former University of Tennessee professor William Sanders and his colleagues, found that the effectiveness of teachers has more of an influence on student achievement than any other schooling factor. They also found a wide range of effectiveness among teachers; there are some very good teachers, some very bad teachers, and a wide range of performance between them. Using the "Sanders methodology," teachers are placed into effectiveness quintiles based on their students' growth in achievement, or the "value added" by the teachers. Those teachers who fall into the first quintile, the least effective teachers, were found to elicit average student gains of roughly 14 percentile points a year. The most effective teachers elicited an average gain of 52 percentile points a year. The effects of teacher quality were also found to persist for years after a student had a particular teacher.

More recently, researchers have sought to isolate teachers' contribution to student performance and assess how much of their overall contribution can be associated with measurable teacher characteristics, such as experience and degree level. Economists Eric Hanushek, John Kain, and Steven Rivkin estimated that, at a minimum, variations in teacher quality account for 7.5 percent of the total variation in student achievement—a much larger share than any other school characteristic.

This estimate is similar to what my colleagues and I found: that 8.5 percent of the variation in student achievement is due to teacher characteristics. We found that the vast majority (about 60 percent) of the differences in student test scores are explained by individual and family background characteristics. All the influences of a school,

Elusive Qualities (Figure 1)

The teacher characteristics that we can measure—experience, education level, certification status, and so on—only explain 3 percent of the differences in student achievement that are attributable to their teachers' influence.



SOURCE: Dan Goldhaber et al., 1999

including school-, teacher-, and class-level variables, both measurable and immeasurable, were found to account for approximately 21 percent of the variation in student achievement. This 21 percent is composed mainly of characteristics that were not directly quantified in the analyses. Since we used statistical models that included many observable school-, teacher-, and class-level variables—such as school and class size, teachers' levels of education and experience, and schools' demographic makeup—it is clear that the things that make schools and teachers effective defy easy measurement.

Only about 3 percent of the contribution teachers made to student learning was associated with teacher experience, degree attained, and other readily observable characteristics. The remaining 97 percent of their contribution was associated with qualities or behaviors that could not be isolated and identified (see Figure 1).

Which Attributes?

What does the empirical evidence have to say about specific characteristics of teachers and their relationship to student achievement? First it is important to note that studies focus on different grade levels, subjects, and types of students taught, and in some cases the estimated effects of particular attributes are not consistent across the board. Furthermore, studies vary considerably in quality. Experiments, which provide the most credible results, are rare in education, and relatively few studies that address students' outcomes observe the professional norm of having detailed controls for students' background characteristics (including previous academic achievement). That said, here is a broad overview of the findings for various teacher attributes.

Teacher degree and experience levels: Teachers' education (degree) and experience levels are probably the most widely studied teacher attributes, both because they are easy to measure and because they are, in the vast majority of school systems, the sole determinants of teachers' salaries. However, there appears to be only weak evidence that these characteristics consistently and positively influence student learning.

In a 1986 literature review (and in follow-up reviews in 1989 and 1997) that has framed much of the debate, Hoover Institution senior fellow Eric Hanushek showed that only a small proportion of studies find these teacher characteristics to be statistically significant in the expected direction.

Not all studies reach the same conclusion. In a 1996 meta-analysis published in the *Review of Educational Research*, Rob Greenwald and his colleagues concluded that "school resources are systematically related to student achievement and that these relations are large enough to be educationally important" and "resource variables that attempt to describe the quality of teachers (teacher ability, teacher education, and teacher experience) show very strong relations with student achievement."

It's true in both sets of reviews that many more studies found statistically significant positive effects for teacher experience than found statistically significant negative effects, but the statistically significant positive findings were found in only about 30 percent of the studies. There are also statistical shortcomings in many of the studies cited by both Hanushek and Greenwald et al.

Part of the explanation for the mixed findings may be that experience and degree level matter only in certain circumstances. For example, there is little evidence that experience beyond the first couple of years in the classroom makes one a better teacher. And teacher experience implicitly captures the effects of the prevalent graduation requirements and labor market conditions at the time when teachers were hired. Furthermore, as I describe below, the effect of degrees appears to hinge on the subjects that are taught and whether the degrees are specific to those subjects.

Subject-matter knowledge: The evidence is somewhat mixed, but it suggests that teachers' knowledge of their subject matter, as measured by degrees, courses, and certification in that area, is associated with high performance. Studies with more detailed measures of teachers' educa-

tion levels and coursework in subject areas found that, at least in math and science, academic preparation does positively influence student achievement. Having an advanced degree in subjects outside of math and science, however, does not appear to affect student achievement. Taking additional courses in one's subject has some effect, but it depends on the level of the courses a teacher is teaching. Perhaps not surprisingly, in a 1994 study David Monk and Jessica King Rice found that the teaching of higher-level courses seems to require greater knowledge of subject matter than does the teaching of lower-level courses.

Teachers' pedagogical knowledge: The value of teaching teachers how to teach, or pedagogy, is more hotly debated. Since there is little research directly assessing the influence of pedagogical training on student outcomes, this debate tends to focus on the impact of teachers' performance on licensure exams and the merits of licensing teachers. A 1986 study found that the average National Teacher Exam scores of all the teachers in a school district have a strong positive correlation with the average performance of students on standardized tests. It is not clear, however, whether this reflects the influence of subject-matter knowledge or pedagogical training, since the National Teacher Exam tests both. A number of studies have found that fully certified teachers influence student achievement positively. For instance, a 1985 review by Carolyn Evertson, Willis Hawley, and Marilyn Zlotnik found that 11 of 13 studies judged regularly certified teachers to be more effective than those who held a provisional or emergency certificate. But only four of these studies were based on students' outcomes and most of them were more than 25 years old, which means they predated the "value added" methodology of assessing educational effects that is now standard practice. More recently, Stanford professor of education Linda Darling-Hammond analyzed state-level data and concluded, "The most consistent highly significant predictor of student achievement in reading and mathematics in each year tested is the proportion of well-qualified teachers in a state: those with full certification and a major in the field they teach." This study, however, potentially suffers from a statistical problem known as aggregation bias. There are many factors that influence achievement at the state level, many of which cannot be identified and controlled for, and it therefore may be inappropriate to conclude that the test-score results are linked to the proportion of teachers with specific characteristics in a state.

Researchers who have studied the Tennessee data found that the EFFECTIVENESS OF TEACHERS has more of an influence on student achievement than any other schooling factor.

Dominic Brewer and I analyzed teacher certification and its relationship to student achievement at the level of individual teachers and found that teachers with standard certification outperform those who are not certified in their field, but they do not appear to be more effective than those who hold emergency credentials (teachers who have not satisfied all of the requirements necessary to obtain a standard certificate). One reason why findings are inconsistent is the lack of a standard system of licensure; each state has a unique set of requirements for individuals who wish to enter the teaching profession, and some states' requirements may be more rigorous than others.

What is to be concluded about the effect of a teacher's level of education and certification status? As Carolyn Evertson and her colleagues write, "Investigations of teacher education do not represent a strong body of research." Most of the studies that find statistically significant relationships between teacher training and student achievement find that the effects of these characteristics are small and specific to certain contexts. If a large, consistent association between teacher training and student achievement existed, it probably would not be all that hard to detect.

Other teacher attributes: Recent studies suggest that measures of teachers' academic skills, such as SAT or ACT scores, tests of verbal ability, or the selectivity of the colleges they attended, may predict their effectiveness more accurately than the characteristics discussed above. However, here too the evidence and the estimated relationships are relatively weak. Greenwald et al.'s 1996 meta-analysis found that teachers' academic skills were shown to have a positive relationship to student achievement in 50 percent of the studies they analyzed, a much higher proportion than for teacher education or experience. Studies by Ronald Ferguson and Helen Ladd found positive relationships between aggregate teacher scores on the ACT, literacy examinations, or states' licensure examinations and aggregate student performance on standardized tests. However, the fact that these studies were done at the aggregate (school or school district) level casts some doubt on them. It's unclear whether higher-scoring teachers lead to higher-scoring students or whether affluent districts, which tend to have higher-achieving students, also tend to hire teachers with higher scores.

In a 1994 study, Ronald Ehrenberg and Dominic Brewer found that students score higher on standardized exams if their teachers attended more selective undergraduate institutions. In a separate study, Ehrenberg and Brewer reexamined the Coleman data and found a significant positive association between teachers' verbal ability and student outcomes. As Hanushek wrote in 1989, "Perhaps the closest thing to a consistent conclusion across studies is the finding that teachers who perform well on verbal ability tests do better in the classroom."

Little Guidance

Good teaching is clearly important to raising student achievement. In fact, most research suggests that the benefit of improving the quality of the nation's teaching workforce is far greater than other policy interventions, such as lowering class size.

However, while we know that good teaching is important, it's far less clear what makes for a good teacher. The measures of teacher quality that are used by most public school systems to screen candidates and determine compensation—certification, experience, and education level—have been well researched, but there is little definitive empirical evidence that these characteristics, defined in general terms, are associated with higher student achievement. Teachers' educational levels appear to make a difference when the education is related to the subject taught, but advanced degrees do not appear to serve as a good measure of quality in general. There is also some evidence that experienced teachers are more effective with students, but the benefits of additional years of experience appear to level off early in a teacher's career. Measures of teachers' academic skills, such as their verbal ability, may more accurately predict their effectiveness, but there is far less evidence on this issue, and these findings are also not conclusive. There is little evidence on the issue of teacher certification as well, and the evidence that does exist is mixed.

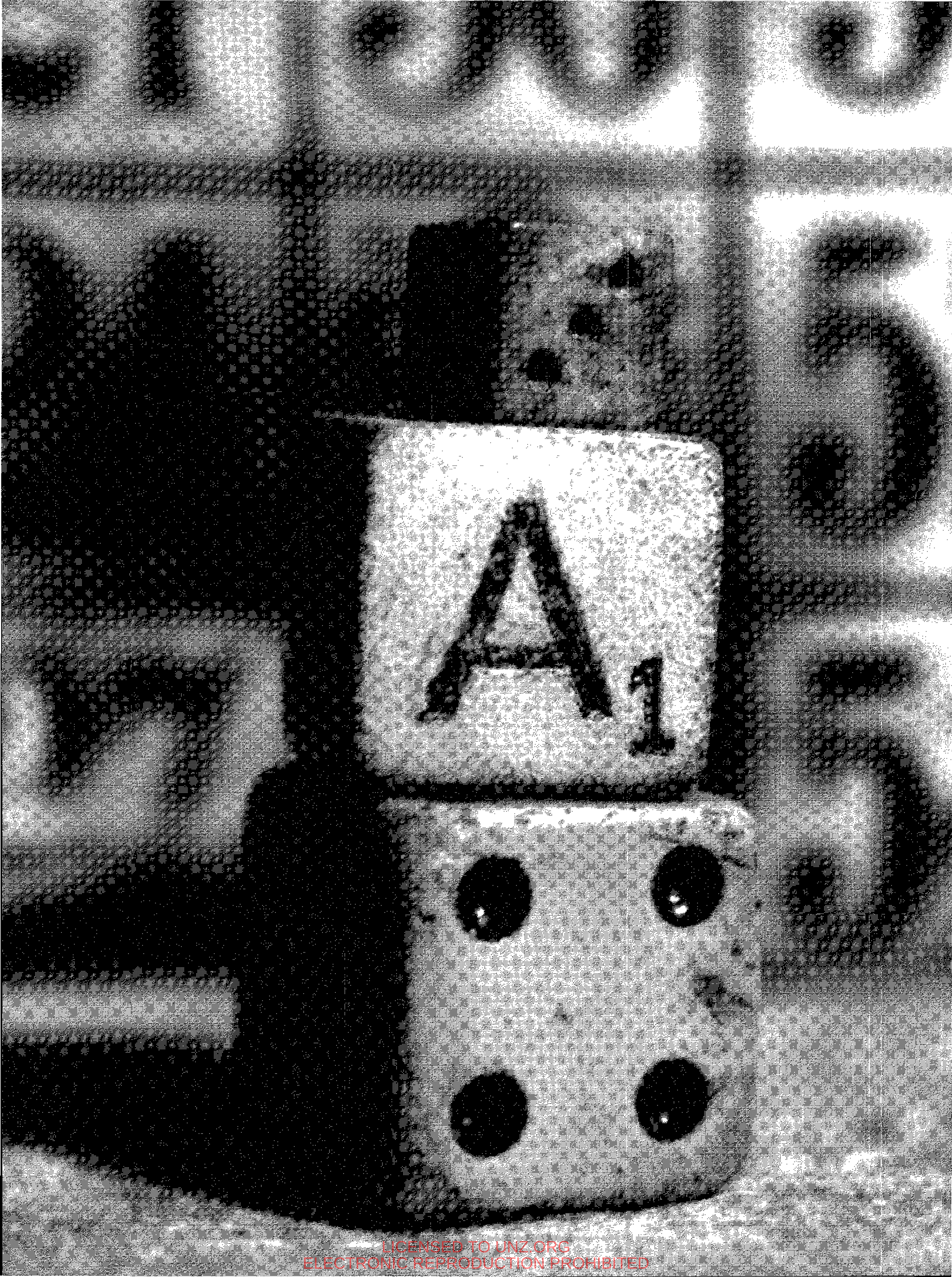
THE LACK OF CLEAR EVIDENCE on the effectiveness of teacher certification suggests that policymakers should continue experiments, such as Teach for America, that allow individuals to enter the teaching profession through alternative routes.

What does the empirical evidence imply for policymaking? First, the importance of teacher quality cannot be overstated. Teachers can have a profound effect on students, and school systems make a significant long-term investment when they hire teachers. Unlike other education investments, such as class size, which may be easily altered from year to year, the tenure system implies that the employment of an individual teacher is near permanent. For these reasons, the selection of teachers is of paramount importance. I would argue that this function of school systems receives too little attention at the local level.

Second, the lack of clear evidence on the effectiveness of teacher certification suggests that policymakers should continue experiments, such as Teach for America, that allow individuals to enter the teaching profession through alternative routes. To the degree that superintendents, principals, and hiring committees accurately identify top-notch candidates through these alternative routes, such programs may serve as an important supply of future teachers. Alternative-route programs should be studied to determine the extent to which state-level restrictions on entry into the market are justified.

Finally, the compensation structure used by virtually all school districts is not well aligned to promote the acquisition of skills found to influence student outcomes. For instance, data from the National Center for Education Statistics show that salary schedules provide pay premiums of about 11 percent for master's degrees and 17 percent for a doctorate. Generally, these premiums are received regardless of whether degrees are specific to the subjects in which teachers teach, despite the evidence that out-of-field degrees contribute little toward student achievement. The compensation structure also does not provide policymakers with tools to address areas of shortage, to reward job performance or the acquisition of skills deemed to be important, or to compensate for the difficulty of a teaching assignment. Do alternatives to the compensation structure currently in place in public schools work? Unfortunately, there is very little evidence to answer this question. However, given the concern about the quality of the existing teacher workforce, experimenting with alternative pay structures may be worthwhile.

–Dan Goldhaber is a senior research associate at the Urban Institute in Washington, D.C.


A1

Failing to account for natural fluctuations in test scores could undermine the very idea of holding schools accountable for their efforts—or lack thereof

by THOMAS J. KANE, DOUGLAS O. STAIGER, AND JEFFREY GEPPERT

randomly Accountable

THE ACCOUNTABILITY DEBATE TENDS TO DEVOLVE INTO A BATTLE between the pro-testing and anti-testing crowds. But when it comes to the design of a school accountability system, the devil is truly in the details. A well-designed accountability plan may go a long way toward giving school personnel the kinds of signals they need to improve performance. However, a poorly designed scheme, which ignores the statistical properties of schools' average test scores, may do more harm than good.

The recent debate over the reauthorization of the federal Elementary and Secondary Education Act (ESEA) is a case in point. From his first days in office, President Bush promised to make education reform a centerpiece of his administration, using the reauthorization of the ESEA as an opportunity to give the state-led accountability movement a dramatic shove forward. Within six months of his taking office, both houses of Congress had passed bills that imposed new federal standards for the states' accountability efforts.

However, both bills were seriously flawed. They created standards that, over time, would have identified nearly every school in the nation as "low performing," forcing them to spend precious resources developing unnecessary school-improvement plans. A tide of paperwork would have crowded out time for learning. This almost turned the most significant federal foray into education policy in decades into an embarrassment. Changes were made by a House-Senate conference committee, so the law, as enacted, remedied the most glaring problems, but created others. The saga illustrates the difficulties of designing an effective accountability system.

PHOTOGRAPH BY TINA WEST/WWW.IMAGES.COM

The House and Senate Bills

At the heart of both bills was a detailed formula for determining when a school is making "adequate yearly progress." The consequences for schools that failed to meet their performance targets were progressively severe—after one year, districts would be required to offer public school choice to all the students in a school; after several years, districts would be required to replace school staff, convert the school into a public charter school, or hand the school over to a private contractor.

Single-year changes in test performance are very unreliable indicators of where a school is headed over the long term.

The problem is that such consequences place too much weight on single-year changes in test scores at the school level. Either bill would have required an increase in the proportion of students scoring above the proficient level in both math and reading, each and every year. However, test scores at the school level often fluctuate for reasons other than any underlying change in a school's performance. Such volatility arises from two sources. The first is variation due to differences in the groups of students being tested each year. Even if the students are being drawn from the same families and the same neighborhoods, the average performance of a school can fluctuate from year to year depending on the attitudes and abilities of the students in each cohort. The average elementary school contains only 68 students per grade level. With a sample this small, having five particularly bright students (or a few students with undiagnosed learning disabilities) in any one year can lead to large fluctuations in a school's test scores from one year to the next. The Department of Labor measures the monthly unemployment rate with a sample of nearly 60,000 households. Congress was proposing that the Department of Education measure the performance of the typical elementary-school grade with a sample nearly 1/1000 the size.

The second source of variation is one-time factors that lead to temporary fluctuations in test performance. Some of these factors are likely to be unrelated to the educational practices of a school. For instance, a dog barking on the day of the test, a severe flu season, or one particularly disruptive student in class could cause scores to fluctuate. There may be other sources of volatility that are more related to the educational mission of a school, such as the favorable chemistry between a teacher and a particular group of students or teacher turnover. Whatever the source of variation, single-year changes in test performance are very unreliable indicators of where a school is headed over the long term.

Consider the examples of North Carolina and Texas. Between 1994 and 1999, these states were the educational envy of the nation, raising proficiency rates in math and reading by 2 to 5 percentage points in the average year. However, the vast majority of schools in those states exhibited much less consistent progress: less than 2 percent of schools witnessed an increase in math and reading proficiency each and every year for those five years. Indeed, we estimate that between 98 and 100 percent of the elementary schools in North Carolina and Texas would have failed the House and Senate's initial definitions of annual yearly progress at least once between 1994 and 1999.

Furthermore, both bills would have compounded the error by requiring annual increases in test scores for every racial subgroup in a school. The intent was admirable: to ensure that schools do not ignore minority children. But this provision was likely to have harmed its intended beneficiaries, by arbitrarily sanctioning schools that enroll students from several different racial or ethnic subgroups. Suppose that a school is solidly on the path to improvement, with a 70 percent chance of increasing the proficiency of any racial subgroup in a given year. A school with two racial subgroups in its student body would have a less than 50-50 chance of achieving an increase for both groups in a given year—because the year-to-year fluctuations are nearly independent for each racial group (therefore the probability is .70 times .70, or .49). The odds would be even longer for a school with three racial subgroups (.70 times .70 times .70, or .34). Since African-American and Latino students are more likely to attend schools with more than one racial group, they are more likely to see their education disrupted arbitrarily.

A number of states have established accountability programs that track the performance of racial and ethnic subgroups separately. For example, California requires schools to meet certain growth targets for all "numerically significant" subgroups in a school. In order to be numerically significant, a group must either represent at least 15 percent of the student body and have more than 30 students or have more than 100 students regardless of what percentage they are. There are eight different groups that can qualify as numerically significant, depending on the number of students in each group in a school: African-American, American Indian (or Alaska Native), Asian, Filipino, Hispanic, Pacific Islander, white non-Hispanic, and "socioeconomically disadvantaged" students.

We calculated the likelihood of a California school's winning a Governor's Performance Award by size and by the number of numerically significant subgroups. Among the smallest quintile of elementary schools, 47 percent of racially heterogeneous schools (those with four or more racial subgroups) won performance awards, versus 82 percent of similarly sized but racially homogeneous schools. This is particularly ironic given the fact that overall growth in performance was slightly higher for more-integrated schools between 1999 and 2000. Moreover, the reason for a school's failure to win an award was often not that African-American and Latino students were lagging behind, but that white non-Hispanic students experienced slower growth in achievement: the average school with multiple racial subgroups witnessed larger gains for African-American and Latino students than for white students.

Separate achievement targets for racial and ethnic subgroups seem to be neither necessary nor especially effective in coaxing schools to focus on the performance of racial and ethnic minorities. In North Carolina, where there are no separate racial targets, African-American and Latino students experienced slightly higher improvements in proficiency than white non-Hispanic youth. Until this year, the rating system in Texas specified separate targets for racial subgroups that accounted for more than 10 percent of the student body (and more than 30 students). However, African-American and Latino students saw the same improvements in their test scores whether or not they attended schools with enough minority students to require a separate racial target.

Remaining Problems

The conference committee's compromise bill remedied some problems but created new ones. Earlier versions of the legislation rated schools according to their year-to-year improvements in the share of their students who achieve a certain proficiency level. Now schools will simply need to have a certain minimum percentage of their students (and of racial, ethnic, and socioeconomic subgroups within each school) deemed "proficient" each year. The initial minimum proficiency rate will be the greater of the proficiency rate of the 20th-percentile school or the average statewide proficiency rate of the lowest-scoring subgroup. In many states, the effective minimum will be the proficiency rate of the 20th-percentile school. However, more than 20 percent of all schools are likely to fail, because the threshold will apply not only to the school as a whole but also to all the subgroups in a school. If any racial, ethnic, or socioeconomic subgroup within a school fails, the school fails. As a result, a disproportionate number of the schools that enroll disadvantaged minority subgroups are likely to fail. The minimum proficiency rate that schools are required to meet will be raised gradually to 100 percent over the next 12 years.

The main beneficiaries of the conference committee's changes will be suburban schools whose initial rates of proficiency are above the minimum, for they will no longer be penalized for temporary downward fluctuations in scores. The primary losers will be schools with initially low levels of proficiency for any subgroup. They will now be required to achieve a 10-percentage-point increase in proficiency for those subgroups to avoid the sanctions in a given year.

One flaw in the new formula is that it provides a strong incentive for states to lower the score students must exceed on their state tests in order to achieve "proficiency." The problem is that redefining proficiency simply because of the new federal requirements may create a credibility problem for the standards movement in a number of states.

Another problem not remedied in the final bill is that any federal definition of adequate yearly progress is likely to conflict with at least one of the state accountability plans that are already in place. There are three common variants in state accountability systems: some states, such as North Carolina, Arizona, and Tennessee, rate their schools with a measure of a school's value-added, using the growth in performance for a given group of students since the end of the preceding school year; other states, such as Texas and Illinois, rate their schools on the percentage of students scoring above certain thresholds; still other states, such as California, rate their schools based on their change in test scores from one year to the next. (A fourth category of states rates schools based on some mixture of value-added, levels, or changes.) Thus states that have been rewarding schools based on value-added measures or on changes in scores may be required to sanction the very schools they have been rewarding.

The next battleground is likely to be the issue of how many students it takes to create a separate racial, ethnic, or socioeconomic subgroup for accountability purposes. The legislation only requires that there be a sufficient number of students to yield statistically reliable information in order for the subgroup to count separately. The higher the threshold—say, requiring a subgroup to represent at least 15 percent of the student body, as opposed to 5 or 10 percent—the lower the failure rate will be for schools with small percentages of disadvantaged minority students.

Design Principles

State and federal officials ought to keep three basic principles in mind in designing test-based accountability systems:

✦ *Multiple years of data are required to measure improvements in performance reliably.*

Children arrive at school with widely varying levels of preparation. Even a mediocre school can expect high test scores if its students come from wealthy backgrounds. As a result, policymakers in many states have attempted to level the playing field by focusing on improvements in test scores. However, improvements are very difficult to discern with a couple of years' worth of data, for two reasons. First, schools differ much less in the extent to which they improve test scores from year to year than they do in their beginning level of performance. Second, any measure of change in performance is likely to amplify the effect of sampling variation and other one-time factors that lead to fluctuations in performance. In other words, identifying improvements in performance as opposed to levels of performance in a single year is like looking for a smaller needle in a bigger haystack. If policymakers intend to measure and reward improvements in test performance at the school level, they will need to rely on multiple years of data.

Improvements can be measured in two basic ways: the improvement in performance for a given group of students from one year to the next (known as a value-added approach), or the improvement in performance across different groups of students (which we will refer to as cross-cohort changes). The improvement of scores for at least two contiguous grades (for example, grades 4 and 5) from one year to the next is a mixture of value-added changes (the 4th grade students who become 5th graders) and cross-cohort changes (the 4th grade students this year are a different group from the 4th grade students last year).

With test scores being so volatile, school personnel are at a substantial risk of being punished or rewarded for results that are beyond their control.

Kane and Staiger have analyzed the statistical properties of value-added and cross-cohort changes in test scores, using data from North Carolina (see Figure 1). (Full citations are available at www.educationnext.org.) We measured value-added with the average change in combined reading and math scores for a school's students between the end of 3rd grade and the end of 4th grade; we measured cross-cohort changes with the change in 4th grade scores from one year to the next. Among median-size schools in North Carolina, roughly half of the variance between schools in value-added in 4th grade math and reading was due to sampling variation and other one-time factors. For the smallest quintile of schools, the percentage of variance due to non-persistent factors was even higher (58 percent), while for the largest quintile of schools the percentage was somewhat lower (29 percent). Cross-cohort changes in mean test scores from one year to the next were measured even more unreliably. More than three-quarters of the variance in the annual change in mean test scores among the smallest quintile of schools was due to one-time, non-persistent factors. This percentage was only slightly smaller (73 percent) for the largest quintile of schools. Such volatility can wreak havoc when rewards and punishments are doled out on the basis of changes in test scores; school personnel are at risk of being punished or rewarded for results that are beyond their control.

Therefore, when policymakers seek to reward schools for improvements in test scores, they should do so based on multiple years rather than a single year of data. Moreover, while a simple arithmetic average of improvements over multiple years would be an improvement, there are even more efficient ways to pool information over time. For instance, building on work by McClellan and Staiger (1999) in rating hospital performance, we have proposed a simple technique for pooling information over time, which improves on a simple arithmetic mean by taking into account the amount of "signal" and "noise" in a given measure of performance. For instance, for large schools, for which we would expect less noise in any given year's measure, the proposed method would place more weight on more recent scores; for small schools, the method would place more equal weights on each of several years' worth of scores.

- *Incentives targeted at schools with test scores at either extreme—rewards for those with very high scores or sanctions for those with very low scores—affect primarily small schools and provide very weak incentives for large schools.*

Each year since 1997, North Carolina has recognized the 25 elementary and middle schools in the state with the highest scores on the "growth composite," a measure reflecting the average gain in performance among students enrolled at a school. Winning schools receive financial awards.

One indicator of the volatility of test scores is the rarity of repeat winners. Between 1997 and 2001, 101 awards were handed out for schools ranking in the top 25. (One year, two schools tied at the cut-off.) These 101 awards were won by 90 different schools, with only 9 schools winning twice and only 1 school winning three times. No school was in the top 25 in all four years.

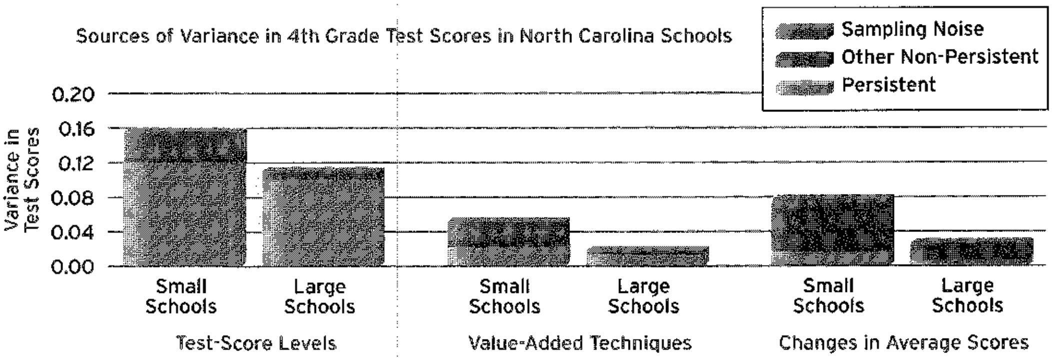
Of the 840 elementary schools we analyzed, 59 were among the top 25 at some point between 1997 and 2000 (the top 25 each year included middle schools, which we are not analyzing here). Among all the schools, the average gain score was not strongly related to school size, but the variance between schools was much larger for small schools. The variance in mean gain scores among schools in the smallest size decile was nearly five times the variance among the largest decile of schools (.048 compared with .011). As a result, schools in the smallest decile were much more likely to be among the top 25 schools at some point over the period: Even though their mean gains were not statistically different, the smallest schools were 23 times more likely to win a top-25 award than the largest schools.

For the very same reason, small schools are also overrepresented among those with extremely low test scores. Beginning in 1997, the state assigned assistance teams to intervene in schools that performed poorly on state tests and failed to meet their growth targets from the previous year. All but one of the elementary schools assigned an assistance team were among the smallest 40 percent of schools. (The smallest decile of schools would have received an even larger share of the assistance teams, except for a rule requiring that the proportion of students scoring below grade level be statistically significantly less than 50 percent.)

This year, the state of California distributed $100 million to teachers in schools that started with test scores in the bottom half of schools in 1999 and achieved large gains in performance between 1999 and 2000. A thousand teachers in schools with the largest improvements received $25,000 bonuses on average.

Value-Added and Change Scores Are Less Reliable (Figure 1)

Of three ways of measuring a school's performance—the level of its test scores, value-added techniques, and changes in the average level—test-score levels are the most reliable. This is because schools differ more in their level of performance than they do in their relative rates of change, and a smaller portion of differences in levels is due to sampling noise or other non-persistent factors, such as a disruption on the day of the test. Reliability problems are most acute for small schools because of the smaller number of observations.



SOURCE: Thomas J. Kane and Douglas O. Staiger, 2001

Small schools in California were considerably more likely to win one of these awards than were larger schools. Given the importance of sampling variation and the fact that the largest bonuses were reserved for teachers in schools with the most extreme increases in test scores, this is hardly a surprise.

A threshold at either extreme is likely to be irrelevant for large schools, since they are unlikely to experience such large swings in performance regardless of their efforts. If the marginal costs of improving are also higher at large schools, the problem of weak incentives for large schools would only be compounded. A remedy would be to establish different thresholds for schools of different sizes. For example, grouping schools according to size (as is done in high-school sports) and giving awards to the top 5 percent in each size class would tend to even out the incentives (and disparities) between large and small schools. An alternative solution would be to establish thresholds closer to the middle of the test-score distribution, where the disparity for large and small schools is less extreme.

Helen Ladd and Charles Clotfelter in 1996 and David Grissmer et al. in 2000 reported evidence suggesting that schools respond to incentives by raising student performance. However, the long-term effects of incentives may be quite different from their short-term effects. Even if teachers are not sufficiently aware of the statistical forces at work to recognize their rather limited influence on test scores in the short run, they may well become aware of this over time. If their best efforts are rewarded with failure one year and less work the following year is rewarded with success, they are likely to form negative opinions regarding the value of their efforts.

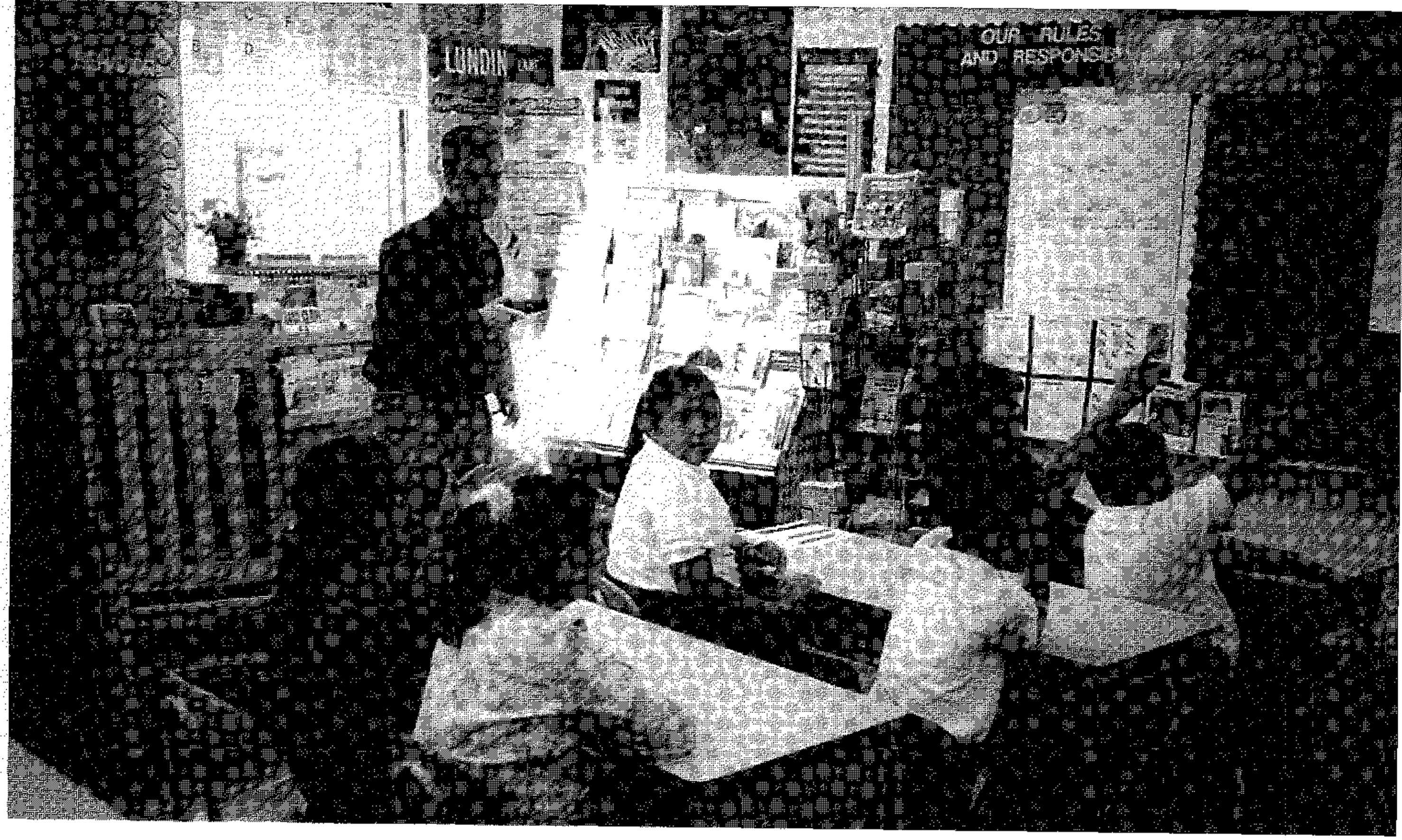
✦ *When evaluating the impact of policies on changes in test scores over time, the natural fluctuations in test scores must be accounted for.*

In 1997, North Carolina identified 15 elementary and middle schools with poor performance in both levels and gains and assigned "assistance teams" of three to five educators to work in these schools. The next year, all of the schools had improved enough to escape being designated "low performing." The state Department of Public Instruction ascribed the improvements to the efforts of the assistance teams; the assistance teams were lauded in *Education Week*'s annual summary of the progress of school reform efforts in the states as well. However, given the amount of sampling variation and other non-persistent fluctuations in test-score levels and gains, schools with particularly low test scores in one year would be expected to bounce back in subsequent years.

The schools that were assigned assistance teams seem to have had a particularly bad year the year they received the sanction. In the year before assignment, such schools had an average 4th grade combined reading and math test score that was .67 student-level standard deviations below the average school. This reveals that they were weak schools the year before being sanctioned. However, in the year of assignment, their average score was even lower, .79 student-level standard deviations below the average school. The year after assignment, their scores seemed to rebound to .52 student standard deviations below the mean. One is likely to greatly overestimate the impact of assistance teams by taking the change in performance in the year after assignment.

There are real differences in performance at the school level. And schools that are not improving should be identified for intervention. However, one year's worth of test-score data is insufficient to discern such differences in a meaningful way. States should be allowed to experiment until the nation finds the ideal way to determine which schools are making adequate yearly progress. We understand the impulse to create a system that requires specific remedies sooner rather than later. However, impatience is an insufficient excuse for bad education policy.

–*Thomas J. Kane is a professor of policy studies and economics at the School of Public Policy and Social Research at the University of California at Los Angeles. Douglas O. Staiger is an associate professor of economics at Dartmouth College. Jeffrey Geppert is a senior research analyst at the National Bureau of Economic Research in Palo Alto, California.*
A portion of this article is drawn from a chapter that will appear in the Brookings Papers on Education Policy 2002 *(Brookings, 2002).*


OUR RULES AND RESPONSIB
LUNDIN

THE Teach for America EVALUATION

Herewith, the first evidence on its recruits' actual performance in the classroom by MARGARET RAYMOND AND STEPHEN FLETCHER

PHOTOGRAPHS COURTESY OF TEACH FOR AMERICA

SINCE 1990 THE NEW YORK–BASED TEACH FOR America (TFA) program has placed more than 7,000 teachers in some of the nation's most challenging school districts. The nonprofit organization recruits high-achieving seniors from top colleges and asks them to commit themselves to two years of teaching in inner-city or rural schools. TFA currently supplies teachers to 18 districts across the country, including Newark, New Jersey; rural districts in the Mississippi Delta; Phoenix, Arizona; and Houston, Texas. Most TFA recruits serve in schools that qualify for funding under Title I of the Elementary and Secondary Education Act due to their high concentrations of students living in poverty. These schools often find it difficult to fill teaching positions with strong candidates.

In number of teachers trained each year, TFA is one of the nation's largest suppliers of teachers. According to the American Association of Colleges for Teacher Education, which represents university- or college-based teacher-preparation programs, only 10 percent of programs offering undergraduate-level training produce more teachers each year. TFA recruits clearly help to alleviate teacher shortages in the school districts in which TFA operates. But how do these recruits perform once they are hired? The evaluation results reported here provide the first evidence of TFA recruits' actual performance in the classroom.

TFA's underlying idea is that many smart, accomplished college graduates would make significant contributions to public education—if it weren't for the fact that they must have majored in education or be willing to spend one to two more years (and significantly more in tuition) earning a master's in education in order to get certified. Even if these bright, young, committed individuals don't remain in teaching forever, TFA posits that they can make a difference in some of the nation's most challenging schools. TFA gives what the program considers potentially excellent teachers—those with both solid knowledge of content and strong leadership skills—a low-cost opportunity to earn a teaching credential while discovering whether they have found a calling. The emphasis on knowledge and leadership is reflected in TFA's recruiting: a typical TFA corps member earned a grade-point average of 3.4 out of 4.0, and 87 percent of recruits have leadership experience.

TFA recruits undergo five weeks of training during the summer. They complete intensive preservice coursework and spend the balance of the summer in classrooms as student teachers or team teachers. Once school starts in the fall, TFA recruits must participate in weekly TFA-sponsored professional development workshops—more than the typical new teacher. This is in addition to any coursework they must complete under their district or state's alternative certification program. In Houston, the TFA site examined for this study, all uncertified teachers must enroll in the district's alternative certification program at the beginning of their first year. It takes a year to earn a credential.

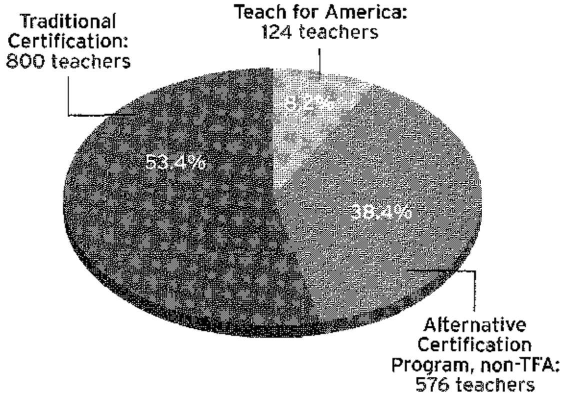
Advocates of alternative certification hail TFA for its creative response to teacher shortages and for its efforts to recruit star college graduates who otherwise might not have found their way to teaching. Many corps members leave teaching after fulfilling their two-year commitment, but they often remain in the education field—as school administrators, in government, or with nonprofits that are involved in education reform. In April 2001 *Education Week* reported, "It's becoming clear that Teach For America is channeling a wealth of talent, energy, and creativity into educational leadership that might otherwise have wound up in such fields as medicine, law, or business."

But TFA has its detractors as well. They point to the lack of pedagogical training and knowledge of child development theories among TFA teachers. They also tend to believe that TFA demeans teaching by treating it as a Peace Corps–style rescue mission rather than a true profession, with salaries appropriate to attracting solid candidates. "A frankly missionary program," wrote Stanford professor Linda Darling-Hammond in an oft-cited 1994 *Phi Delta Kappan* article, "TFA has recruiters and advocates who have focused much of their attention on the advantaged college graduates for whom TFA serves as something useful to do on their way to their 'real jobs' in law, medicine, or business."

Another Source of New Teachers (Figure 1)

In the 2001–02 school year, Teach for America supplied 8 percent of the Houston school district's new hires.

Sources of Houston's Teachers, 2001–02



SOURCES: Teach for America; Houston Independent School District

Into the Breach

So far, these battle lines have been drawn with only rhetoric as a weapon, for evidence on the performance of TFA teachers in the classroom has been nonexistent. To help fill this vacuum, CREDO (formerly known as the Center for Research on Education Outcomes), an independent nonpartisan research group at the Hoover Institution of Stanford University, evaluated the performance of TFA teachers relative to both new teachers and all teachers in the Houston Independent School District. Houston is the 7th-largest school district in the United States, encompassing 186 elementary schools and 34 middle schools. Houston hires approximately 350 new elementary- and middle-school teachers a year; since TFA began supplying teachers to the district, its proportion of new teachers has ranged between 5 and 10 percent (see Figure 1). Latino (54 percent) and African-American (33 percent) students make up the majority of the district's student population. The majority of students live in relatively poor households, with 75 percent receiving free or reduced-cost lunches.

To perform the analysis, the year-end learning gains of students with TFA teachers were measured against the gains of students with non-TFA teachers. Individual scores on the annual Texas Assessment of Academic Skills (TAAS) in mathematics and reading and English language arts were used as the measure of student performance. Texas has annually tested students in grades 3 through 8 since 1993. Having multiple years of test scores for the same student allowed us to estimate the annual progress of individual students by controlling for their previous year's TAAS score.

Many corps members leave teaching after fulfilling their two-year commitment, but they often remain in the education field.

Linked student- and teacher-level data were obtained through the Texas Schools Project at the University of Texas at Dallas. Two groups of students were studied: elementary-school students in grades 3 through 5 (where most students have a single teacher throughout the year) and middle schoolers in grades 6 through 8 (where students tend to have a different teacher in each subject). This article presents the findings for the elementary grades only. The analysis of middle-school teachers produced similar results, but faced methodological challenges arising from students' having had multiple teachers. (For a description of the middle-school analysis, see the full report at: http://credo.stanford.edu/working_papers.htm.)

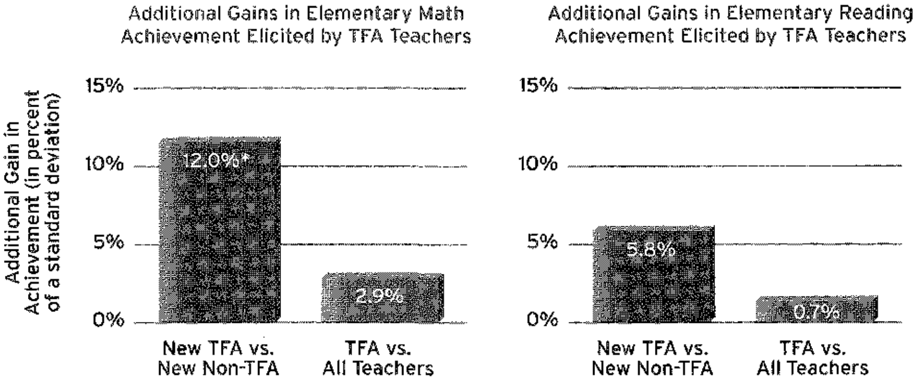
Our study was framed around two separate but related questions. First, we wanted to know whether TFA recruits were effective teachers. So we compared the performance of all TFA teachers with that of all non-TFA teachers in Houston, regardless of how experienced they were. The relevant comparison group was all Houston teachers in grades 3 through 5 from 1996 to 2000.

The second analysis was focused on the question that districts must ask when considering whether to work with TFA: How do TFA recruits measure up against other potential new hires? In other words, if a district had one remaining position to fill, would it be better off with a new TFA teacher or another candidate? For this analysis, we compared TFA teachers with one or two years of experience with other teachers in the district with the same amount of experience.

It is important to note that the group of new teachers with which TFA teachers were compared included teachers with other forms of alternative certification along with a number who were not fully certified. Such teachers need to be included because the purpose of this study is to compare TFA teachers with the pool of potential hires available to the district. As noted above, in both the comparisons of TFA teachers with all teachers and of TFA teachers with only new teachers, TFA teachers represented

Students of Teach for America Recruits Learn More (Figure 2)

New TFA recruits elicited gains in student achievement that were 12 percent of a standard deviation higher in elementary-school math and 5.8 percent higher in reading than those of students of other new teachers.



Additional Gains in Elementary Math Achievement Elicited by TFA Teachers

Additional Gains in Elementary Reading Achievement Elicited by TFA Teachers

***Denotes statistical significance at $p < .05$**

SOURCE: Authors' calculations

a small proportion of the total number of teachers. Achieving statistically significant results requires very large impacts under such circumstances, and so our findings, though stable, do not always reach a level that passes the conventional significance threshhold.

Our study examined the performance of TFA and non-TFA teachers from two perspectives. First, we made a straightforward comparison of the average test-score gains in classrooms run by TFA and non-TFA teachers, controlling for a variety of factors known to influence academic achievement, including students' backgrounds, the students' previous performance on the TAAS, characteristics of their schools, and characteristics of their classmates. In this way, we could isolate the typical effect of having a TFA teacher versus a non-TFA teacher. Second, we wanted to examine the distribution of performance among TFA and non-TFA teachers. For instance, how do high-performing TFA teachers compare with high-performing non-TFA teachers? How bad are the worst TFA teachers? The worst non-TFA teachers? Do the TFA teachers vary much in quality, or are they fairly similar? To address these questions, we estimated how much of the achievement gains in each classroom was attributable to that classroom's teacher.

Our study analyzed and controlled for the following factors:

Teacher Characteristics

- TFA status
- Years of teaching experience

Student Characteristics

- The school attended
- School year
- Minority status
- Eligibility for free or reduced-cost lunch (a proxy for low socioeconomic status)
- Date of birth
- Assigned teacher
- English language proficiency
- Testing exemption status
- Test scores on the annual TAAS exam in each year in which the student was enrolled in a Texas school

School Characteristics

- Share of students who are African-American
- Share of students who are Latino
- Share of students receiving free or reduced-cost lunch

Classmate Characteristics

- Share of a class receiving free or reduced-cost lunch
- Share of a class that scored below the state mean on the previous year's TAAS exam

Findings

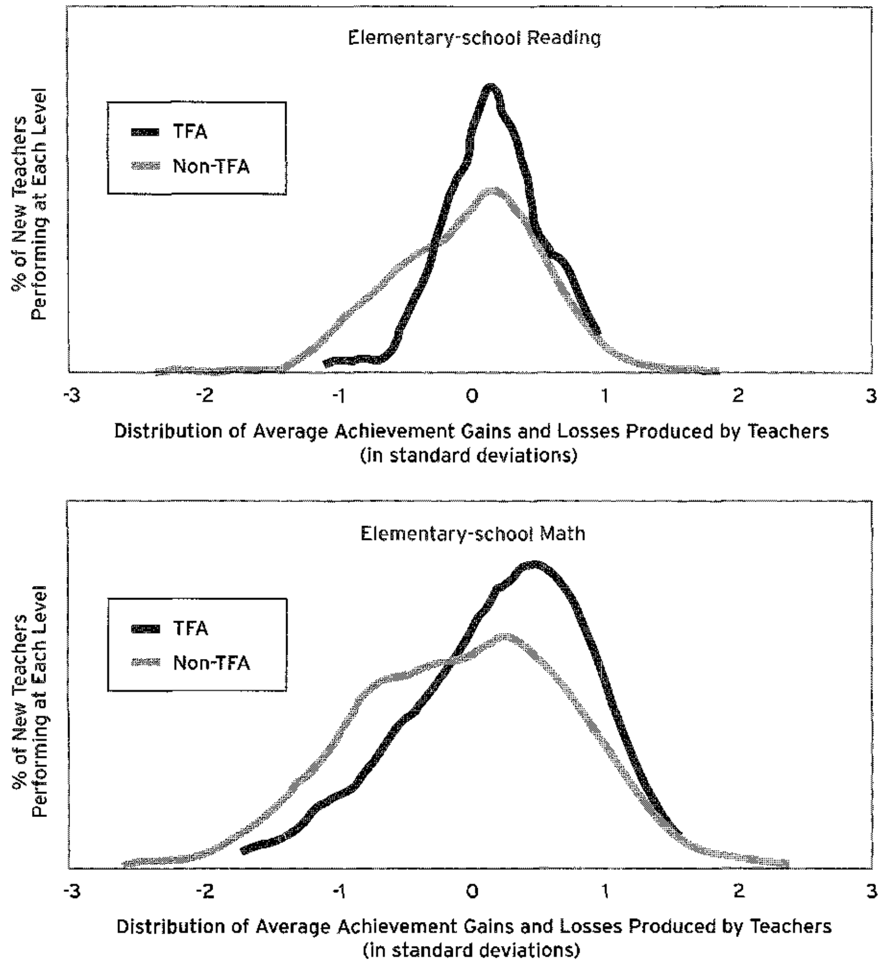
Before the analysis was concluded, students' raw scores on the TAAS exam were standardized to a mean of 0 and a standard deviation of 1. This procedure was performed separately for each year in order to account for variations in the TAAS test from year to year. The average score for a given year was transposed to 0, and all the other scores were distributed around that point in a standard normal distribution. Gains are therefore expressed in terms of standard deviations. A gain of 5 percent of a standard deviation in one year may be considered large, if it is thought that students will continue to experience similar annual gains through the elementary and secondary years. (A full standard deviation is approximately the difference between the test scores of black and white students nationally. More than half of the black-white test-score gap could be erased by steady gains of 5 percent a year among black students.)

Reading. In the elementary grades 3 through 5, students of new Teach for America teachers gained an average of 5.8 percent of a standard deviation more on the TAAS reading exam than did students with other new teachers, a difference that fell just short of statistical significance (see Figure 2). The students of the entire sample of Teach for America teachers made gains that were essentially similar to those made by students of all teachers in the Houston district.

There tended to be less variation in the performance of TFA teachers than there was among non-TFA teachers.

Better and More Consistent (Figure 3)

Of new Teach for America recruits, 63 percent elicited achievement gains in elementary-school reading that were higher than the median new non-TFA teacher was able to produce. The share was 64 percent in elementary-school math. The performance of TFA recruits was also more bunched up around the median, meaning that they were also more consistent in their results, making them less risky to hire.



SOURCE: Authors' calculations

More than 63 percent of TFA teachers generated achievement gains in reading that were higher than the median achievement gains for new non-TFA teachers (see Figure 3). More than 60 percent of TFA teachers did better than the median performance of all teachers. Moreover, TFA teachers tended to be more bunched up around the median than their non-TFA peers. In other words, there were fewer extremely low- and high-performing teachers among the Teach for America recruits; they were more consistent in eliciting achievement gains. The difference between the two distributions was found to be statistically significant in both cases.

Mathematics. Elementary students benefited from having a TFA teacher in math as well. The achievement gains of students with new TFA teachers were 12 percent of a standard deviation higher than those with other new teachers, a result that was statistically significant (see Figure 2). Students of all TFA teachers gained 2.9 percent of a standard deviation more in math than did students of all teachers in the Houston district, a difference that was not statistically significant.

Let's look again at the distribution of performance among TFA and non-TFA teachers. Of new TFA teachers, 64 percent generated achievement gains among their students that were higher than the median achievement gains for new non-TFA teachers (see Figure 3). Fifty-eight percent of TFA teachers performed above the median for all teachers. The lowest-performing new teachers were much worse than the lowest-performing TFA teacher. The best-performing TFA teachers were better than virtually all the teachers in the district. In other words, if you were choosing between two math teachers, and the only thing you knew about them was that one was a TFA member and one was not, you would choose the TFA member. This would give you the best chance of selecting a good teacher.

Our results also suggest that years of teaching experience significantly affect student performance. The largest gains in teacher effectiveness occur early in a teaching career and diminish thereafter until flattening out at around 8 years for math and 11 years for reading. The effect of experience operates independently of teachers' characteristics, so by isolating these effects, we provided a truer test of differences in teachers' backgrounds.

More and Better

TFA appears to be a viable alternative source of high-quality teachers, compared with both the pool of new teachers available for hire in Houston and the district's entire workforce of current teachers. All our results show the average TFA teacher improving her students' performance by more than new teachers and at least as much as all teachers in Houston. The advantage of having a TFA teacher was largest in math, though TFA teachers also generated gains stronger than those of non-TFA teachers in reading. This dispels the notion that TFA is inferior to other sources of teachers. The findings in reading and the comparison between TFA teachers and all teachers in math were not statistically significant, owing largely to the small numbers of TFA teachers in proportion to the respective comparison groups. However, the fact that across many comparisons all the findings pointed in one direction is a strong indication that the general pattern is correct: that TFA teachers produce results as strong as or better than their peers.

Of course, as with any program, there were some TFA recruits who did not perform well in the classroom. This is likely to continue. However, the distribution of performance among teachers clearly shows that the lowest-performing teachers were consistently not TFA teachers and that TFA teachers make up a disproportionate number of high performers. Moreover, TFA teachers were more consistent in their performance than their non-TFA peers, which means that schools face lower risks in hiring them. The fact that many of the highest-performing teachers in the entire district were supplied by TFA is especially meaningful in light of the program's small contribution to the size of the district's workforce compared with other sources of new teachers.

Several caveats are in order. First, this study was based on data from a highly dynamic time in the Houston school district—and in public education more generally. Schools and teachers in Houston were subject to a wide variety of other programs, constraints, and opportunities that no analysis could possibly capture.

Teach for America teachers make up a disproportionate number of the high-performing teachers in the Houston Independent School district.

Furthermore, these findings cannot tell us what aspects of TFA and non-TFA teachers accounted for the differences in their students' performance. Is the effectiveness of TFA teachers due to the type of people being recruited, the difference in their academic backgrounds, the support provided by TFA, participation in the district's alternative-certification program, or a combination of factors? We encourage further study of these questions.

A third caveat relates to the criticisms levied against us by the National Commission on Teaching & America's Future (NCTAF) after our report's initial release. NCTAF, a strong promoter of traditional teacher-education programs and more stringent certification requirements, writes, "[CREDO] does not present data indicating how TFA teachers performed compared to teachers who came into the profession fully qualified and certified to teach. . . . To say that TFA teachers do just about as well as other new teachers [in Houston], an extraordinary number of whom are extraordinarily underqualified, is a weak endorsement at best."

First, this study was not designed to test the benefits of the various teacher-preparation programs or of certification. We do not have the ability, for example, to isolate the effect of Houston's alternative certification program, because an appropriate comparison group was not available. Instead, our evaluation was explicitly intended to assess Teach for America relative to all other sources of new teachers currently available to school districts like Houston. The Houston district draws from a pool of fully certified, traditionally prepared teachers and those who agree to work toward certification through the district's alternative certification program. Our results show that TFA teachers perform modestly better

If you were choosing between two math teachers, and the only thing you knew about them was that one was a TFA member and one was not, you would choose the TFA member.

than those from other sources of new teachers. Studies of the relative merits of various certification policies and programs are clearly important to the continuing policy debates concerning teacher preparation and teacher quality, but they ask questions that are fundamentally different from those of the study described here.

Second, we didn't just compare TFA teachers with the Houston district's other new hires, a fair share of whom are uncertified and didn't attend a traditional education school (though not nearly as many as NCTAF claims). We also compared TFA teachers with all the teachers in the district, the vast majority of whom are certified. The TFA recruits performed just as well as—if not better than—this pool of teachers; in fact, at least in elementary-school math, the best TFA teachers performed better than the district's best teachers after the data were adjusted for the effects of teaching experience.

Informing Policy

While these findings cannot speak to the effectiveness of various certification policies, they at least dispel the notion that only the traditional route to teaching can produce good teachers. What needs to be studied is the degree to which subject-matter skills and strong performance in college can substitute for pedagogical training. TFA corps members are an admittedly select group of college graduates, culled from the finest universities and often performing near the top of their class. These are not the types of students who ordinarily go into teaching. It's possible that traditional certification programs and pedagogical training are less necessary for them than they are for the typical teacher.

The implications of the findings extend beyond teaching preparation. Earlier research has identified strong returns to experience in the early years of teachers' careers. The year-to-year growth in teachers' effectiveness, as measured by student achievement, is largest in the early years of their careers and then quickly tapers off. In the past, attention has been paid to the average student gains for each year of teaching. The findings in this study suggest that the benefits of teaching experience may not be uniform—that new teachers may vary considerably in their initial starting points and in how quickly they improve. Therefore, differences in the value teachers glean from additional professional development may also be in play. These questions cannot be answered without further study. The point is raised here, however, as a reminder that policy decisions concerning teacher training and certification cannot assume that a fixed set of requirements will make all teachers perform the same.

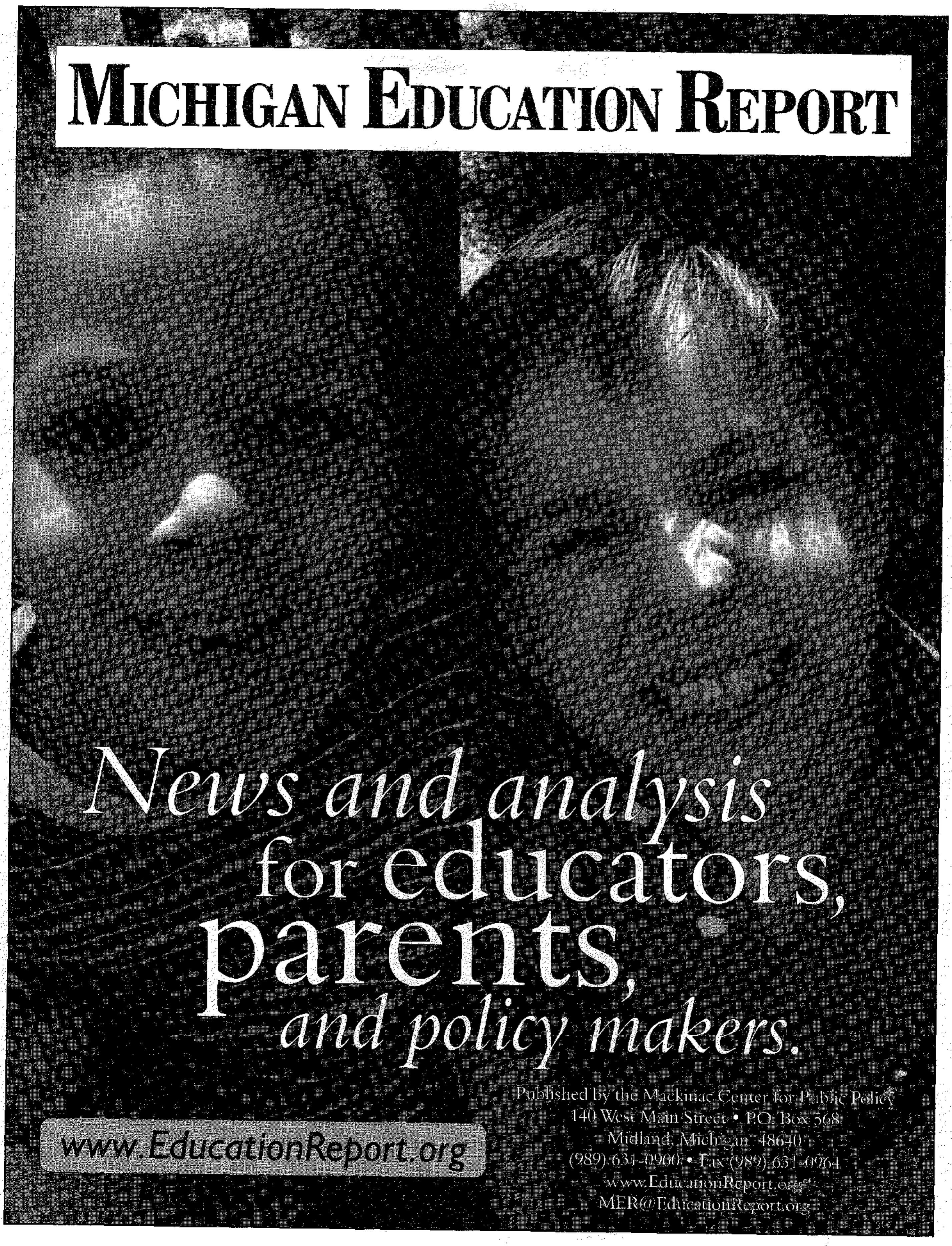
What needs to be studied is the degree to which subject-matter skills and strong performance in college can substitute for pedagogical training.

The present analysis of teacher effectiveness raises other questions beyond certification. One cherished tenet of the NCTAF position is that TFA is pernicious because the majority of TFA teachers leave after completing their two-year commitment. However, the mythical teacher-for-life doesn't exist. Across the four cohorts of teachers included in this study who completed two years or more of teaching, more than a third of TFA teachers remained past their commitment. The NCTAF criticism ignores the fact that 45 to 50 percent of other new teachers leave after two years. The NCTAF assertion that no benefits can accrue to either students or the district from teachers who leave after a few years must apply to both groups or not at all. On the contrary, the evidence here points to the opportunity to identify early in teachers' careers whether they are effective, thus allowing those who are not effective to be encouraged to consider alternative occupations.

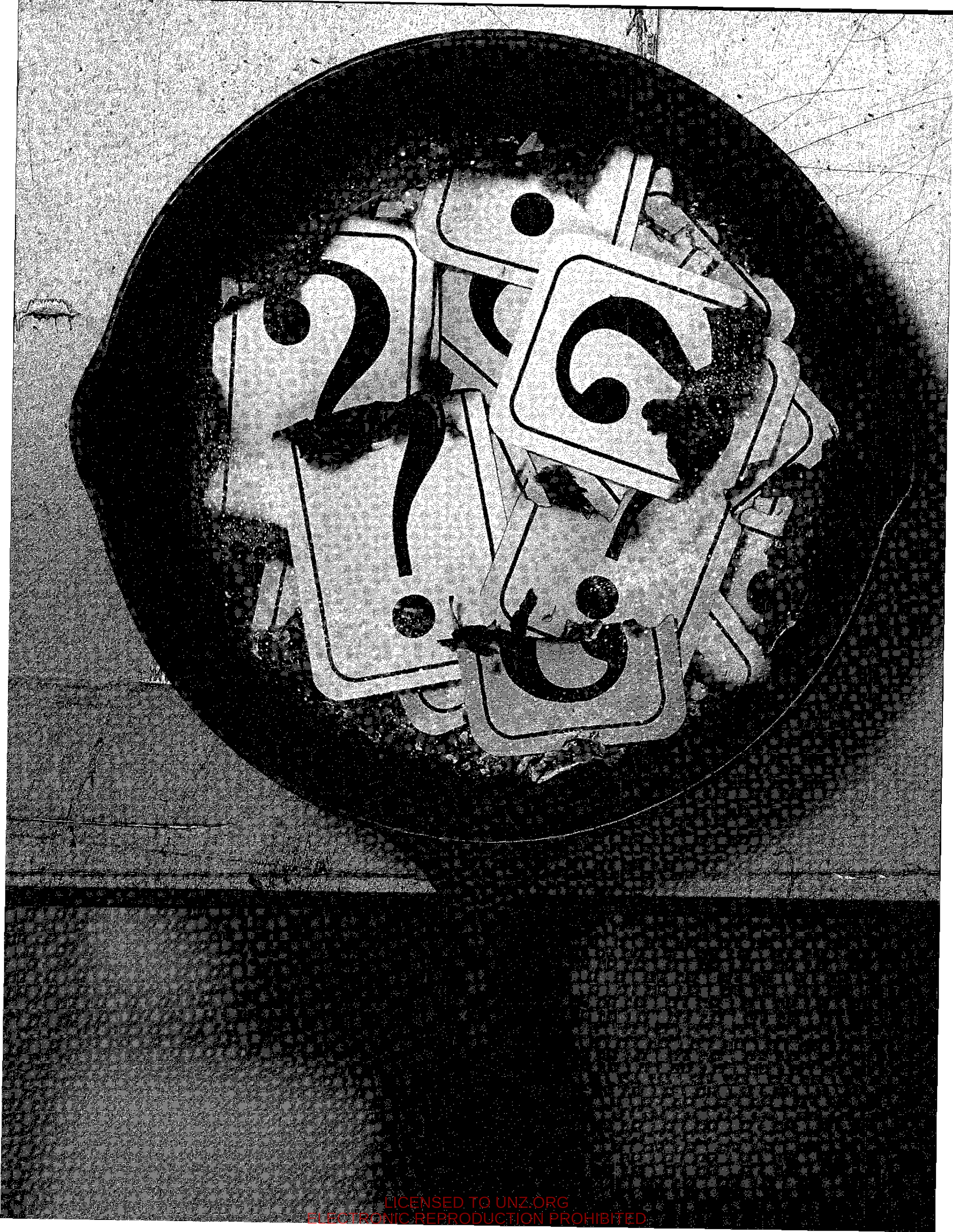
Once other states begin to develop the rich data resources found in Texas, this evaluation should be expanded to include other TFA districts. Corroborating and expanding the results obtained here would contribute to both the TFA program itself and the larger policy world in which TFA continues to grow. TFA has been shown to be a viable source of new teachers for Houston, both in number and in quality. With continuing attention to retention during their two years of service and with many TFA teachers choosing to remain after their commitment has been fulfilled, TFA is likely to create an enduring positive presence in the Houston Independent School District and elsewhere.

Expanding the supply of high-quality teachers is of the utmost importance, but there is little consensus on how to do so. The findings reported here suggest that multiple paths into teaching are both possible and desirable. They suggest that the way to ensure high-quality teaching is to set performance goals and hold teachers accountable for meeting them. The ultimate measure of a teacher-preparation program, traditional or alternative, is its ability to produce good teachers, not its ability to meet an unrelated set of requirements that are not grounded in empirical proof of student performance.

–Margaret Raymond is the director of CREDO, a nonpartisan research group at the Hoover Institution, Stanford University. Stephen Fletcher is the assistant director of CREDO.


MICHIGAN EDUCATION REPORT
News and analysis
for educators,
parents,
and policy makers.
www.EducationReport.org

Published by the Mackinac Center for Public Policy
140 West Main Street • P.O. Box 568
Midland, Michigan 48640
(989) 631-0900 • Fax (989) 631-0964
www.EducationReport.org
MER@EducationReport.org



check the facts

by TERRY M. MOE

Phi Delta Kappa's annual poll is regarded as the definitive measure of where Americans stand on education issues. But are its surveys biased and its reporting self-interested?

COOKING the questions

The 33rd Annual Phi Delta Kappa/ Gallup Poll of the Public's Attitudes Toward the Public Schools

By Lowell C. Rose & Alec M. Gallup

Phi Delta Kappa International, 2001.

Support for vouchers is declining. Or so we are told by Phi Delta Kappa (PDK), the nation's premier association of education professionals, whose annual poll (conducted by Gallup) is widely accepted as the definitive measure of where Americans stand on education issues.

On the surface, it might seem that PDK is just documenting the obvious. After all, two voucher initiatives—one in California, the other in Michigan—were roundly defeated in the 2000 elections, and since then the buzz in education circles is that vouchers have dropped in popularity.

Much of this talk, however, is part of a public relations campaign being waged by the opponents of vouchers, whose aim is to persuade policymakers to stay away from the issue. Such a campaign is to be expected from the teacher unions and their allies, because this is the way the game of politics is played. My fear, though, is that PDK is actively participating in this spin campaign and has been for years.

I cannot read the minds of PDK's researchers, and I do not want to accuse them of such a thing. But I'm convinced that one of two conclusions is justified. Either PDK's polls have purposely been designed to reflect negatively on the voucher issue. Or its researchers have been careless in their design decisions—which I doubt.

PDK's Key Measure

From the 1970s until 1991, PDK measured voucher support with a survey item that defined vouchers as a government-funded program allowing parents to choose among public, private, and parochial schools. After support rose to 50 percent (with 39 percent opposed) in 1991, PDK abruptly dropped this item in favor of a new one. The new question read: "Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?" This question, first asked in 1993, gave results that were strikingly more negative: only 24 percent expressed support (see Figure 1). Indeed, it indicated that even private school parents were

ILLUSTRATION BY JOHN WEBER

opposed to vouchers, a result no expert would be prepared to believe.

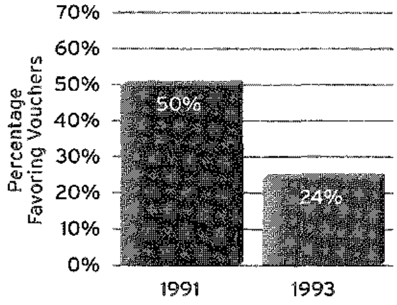
Why such different "facts"? Research has long shown that most Americans are poorly informed about public policy and don't have well-developed views on most issues. Recent polls have shown the same for vouchers. This does not mean that Americans can't connect the voucher issue to their own values and beliefs. But it does mean that, because they come to any survey with little information, they will be quite sensitive to information contained within the survey itself, especially to the specific wording of questions and the order in which they are asked. This information determines how the issue is "framed." And the framing, in turn, influences which (of many possible) values and beliefs get activated in people's minds, and thus how people respond.

All public-opinion researchers are well aware of this. And they all know that, if public opinion is to be well measured on an issue, the issue must be framed with great care. The framing should provide respondents with enough information to give them a good sense of what the issue is about. The information also needs to be balanced, so that respondents are not pushed to see the issue in a positive or a negative light.

PDK's "at public expense" item does not even come close to meeting these basic criteria. The central purpose of a voucher program is to expand the choices available to all qualifying parents, especially those who now have kids in public schools. But the PDK item does absolutely nothing to convey this information. It says nothing about choice, nothing about public school parents' being eligible to participate. Instead, it focuses entirely on private school parents and asks respondents whether the government ought to be subsidizing them. Vouchers are presented, in effect, as a special-interest program for an exclusive group.

Engineered Results? (Figure 1)

The public's support for vouchers appeared to plummet after Phi Delta Kappa altered its voucher question in a way that was sure to elicit negative responses.



1991 PDK survey question: In some nations, the government allots a certain amount of money for each child's education. The parents can then send the child to any public, parochial or private school they choose. This is called the "voucher system." Would you like to see such an idea adopted in this country?

1993 PDK survey question: Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?

SOURCE: Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools, 1991 and 1993

Phi Delta Kappa's question presents vouchers, in effect, as a special-interest program for an exclusive group of private school parents.

This is bad enough. But PDK's researchers compound the bias with the way they choose to inform survey respondents that vouchers are funded by the government. They could have said just that, or they could have found some other neutral way of wording it. Instead they settled on the phrase "at public expense"—which is implicitly pejorative and begs for a negative reaction.

By scientific standards, PDK's "at public expense" question is a poor measure of voucher support. It should never have seen the light of day. Nevertheless, PDK not only adopted this item as its own, but has persisted in using it in every annual survey but one since 1993—with results that, predictably, are on the low end of what we would expect, given the results of better-worded polls on the subject.

PDK's Second Measure

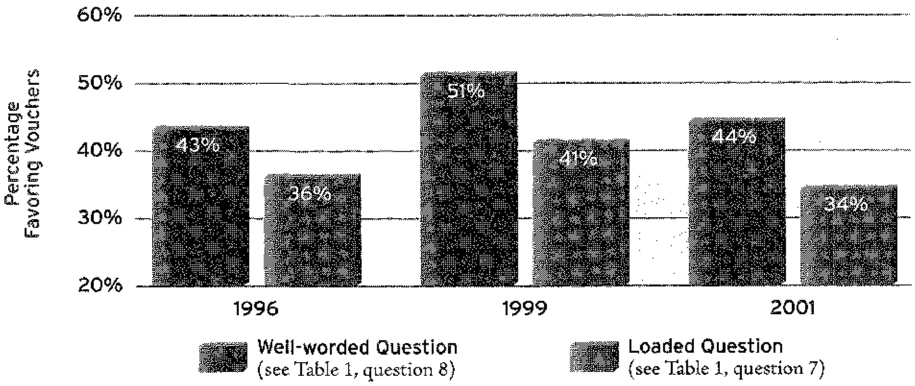
Interestingly, soon after the "at public expense" item made its appearance, PDK introduced a second question to measure support for vouchers. This one is actually informative and neutral, precisely the kind of item that should have been used all along. As such, it gives the appearance that PDK is seriously trying to get a valid measure of voucher support. It reads, "A proposal has been made that would allow parents to send their school-age children to any public, private, or church-related school they choose. For those parents choosing nonpublic schools, the government would pay all or part of the tuition. Would you favor or oppose this proposal in your state?"

Year after year, the results from this second item always show higher support for vouchers than the "at public expense" item does. In 2001, for instance, the "at public expense" item produced a support level of 34 percent, while the second item showed support to be 44 percent. In 1999 the comparable figures were 41 percent and 51 percent, and in 1996 they were 36 percent and 43 percent (see Figure 2).

These higher scores from the second support item, however, do not find their way into PDK's press releases. The media annually turn to PDK as the official source of data on how Americans view the voucher issue, and PDK provides

Wording Effects (Figure 2)

Support for vouchers is always higher in Phi Delta Kappa's surveys when a well-worded question is posed. Yet PDK's press releases report only the results from the survey's loaded question.



SOURCE: Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools, 1996, 1999, and 2001

the lower scores from its "at public expense" question. These are the results that show up in the newspapers, on TV, and on the desks of policymakers.

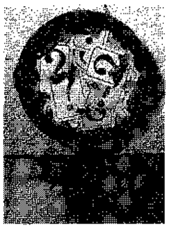
For readers who want to dig deeper, the higher scores from the well-worded item can be extracted from PDK's longer published report. But even these more believable scores must be interpreted with caution. The reason is that when the survey is actually administered to respondents, PDK always asks the well-worded question immediately *after* the "at public expense" question. This means that the voucher issue is already negatively framed before respondents get to the well-worded question, so its scores are also likely to be biased downward.

It doesn't take a rocket scientist to see that there is a problem here. The effects of question ordering are well understood, and competent researchers always think carefully about where each question is placed. It is hard to believe that PDK's researchers are somehow oblivious to all this and are suppressing the higher scores by accident.

The Gallup Experiment

In January 2001, researchers at Gallup carried out an experiment that was written up and posted on the Internet. Noting that measured support for vouchers seems to vary with question wording, they took the two PDK items as their models and submitted a version of each to separate samples of respondents. Would these items produce different levels of support, and how different would they be?

This exercise was not as enlightening as it could have been. The Gallup researchers did not retain the identical wording of each PDK item, but instead tried to improve the questions by tinkering with words they thought might be sources of bias. In the process, the phrase "at public expense" was actually eliminated from the "at public expense" item—an obvious improvement, but one that compromises the experiment.



By scientific standards, Phi Delta Kappa's "at public expense" question is a poor measure of voucher support. It should never have seen the light of day.

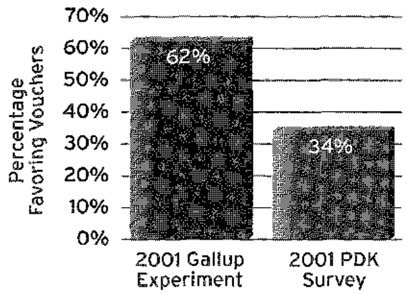
The findings, however, are still eye-opening. The new version of the "at public expense" question asked, "Would you vote for or against a system giving parents government-funded school vouchers to pay for tuition at a private school?" This wording at least retained its focus on private school parents. The result: 48 percent in favor, 47 percent against. The second PDK item became the following: "Would you vote for or against a system giving parents the option of using government-funded school vouchers to pay for tuition at the public, private, or religious school of their choice?" This item suggests that the purpose of the program is to expand choices for a broader population of parents. The result: 62 percent in favor.

A few comments. First, PDK's claim that Americans are turned off by vouchers is simply untrue. When the purpose of a voucher program is well conveyed, most Americans respond positively. In this case, by a very big margin. Second, if this experiment is any indication, the well-worded item on PDK's own survey *is* downwardly biased, despite its reasonable wording. When it is not placed immediately after the "at public expense" question, it produces higher support scores. Third, Gallup's well-worded question produces a support score that is 14 percentage points higher than the special-interest item's score. The gap would likely have been bigger still if Gallup had retained the pejorative phrase "at public expense." A reasonable measure of voucher support, fairly tested, gives *much* higher support scores than the "at public expense" item does (see Figure 3).

The Gallup experiment was available to PDK researchers well before they conducted their 2001 survey. One would think that, in light of this information, objective researchers would have modified their survey, or at least discussed the problems of measurement and interpretation that the Gallup experiment clearly raises. But nothing like this occurred. They designed their items as they always had, and when the results were in, they presented the "at public expense" findings to the media as hard evidence that Americans don't like vouchers.

PDK Cuts Support for Vouchers in Half (Figure 3)

Phi Delta Kappa loudly trumpeted the results from its loaded and poorly framed "at public expense" question. But the results of Gallup's experiment show that a well-worded and unframed question yields support that is much higher.



Gallup experiment question: Would you vote for or against a system giving parents the option of using government-funded school vouchers to pay for tuition at the public, private, or religious school of their choice?

PDK question: Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?

SOURCES: 2001 Gallup Poll Experiment on Question Order and Wording; Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools, 2001

Other Voucher Surveys

Over the years, many organizations have carried out surveys that ask questions about vouchers. Their surveys just haven't received as much attention as PDK's have.

It would be nice if these studies could somehow yield a single, coherent perspective on the voucher issue, but comparing them with any precision is a tricky business. Each has its own voucher item, its own ordering of questions, and its own range of topics being covered (which usually go well beyond education). All of these differences are likely to influence the results on voucher support. When these influences are combined with chance fluctuations due to sampling error—inherent in all surveys, regardless of how carefully they are designed—it is often impossible to tell exactly why surveys yield different results.

Even so, there is helpful information here. An important pattern in these studies is that voucher questions usually come in two types, which mirror the types we've been discussing. The first focuses attention on government subsidies for private school parents. The only real departure from PDK's approach is that government funding is described in some neutral way, without the pejorative "at public expense"—a phrase that no one but PDK is inclined to use. The second type is worded to suggest that vouchers would expand choices for parents generally and that parents with children in public schools would be part of the program.

Not surprisingly, questions of the second type tend to produce much higher support scores than items of the first type do. Furthermore, they show that a majority of Americans tend to express support for the central purpose of a voucher program. The scores jump around from study to study, for all the reasons I've noted, and we are wise not to read too much meaning into particular findings or comparisons. On average, though, the existing studies tend to confirm—many times over—what we already know based on the PDK and Gallup results.

The public's support for vouchers rose to 62 percent when Gallup used a question that was neutrally worded. This should have at least caused Phi Delta Kappa to rethink its techniques.

One issue needs addressing, however. Rather often, the survey questions used by these other organizations have been of the first type, which focus on government subsidies to private school parents. The researchers behind these surveys are presumably objective and competent. Why would they word their questions in ways that fail to convey the central purpose of a voucher program?

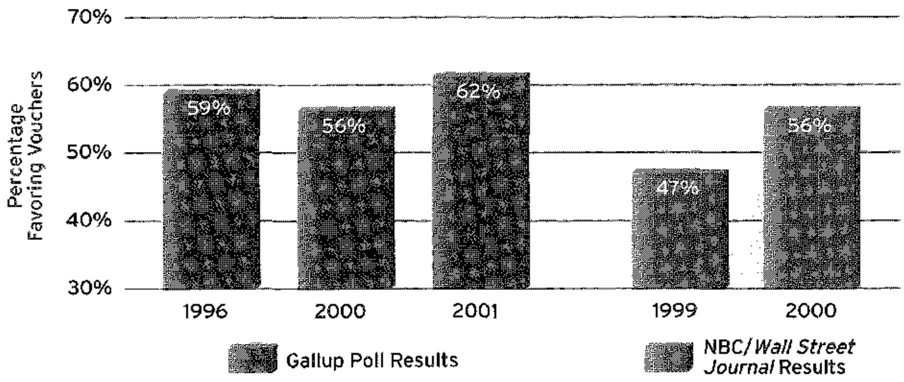
Again, I can't read their minds, but here is my best guess. Sponsoring groups like CBS, NBC/*Wall Street Journal*, and ABC/*Washington Post* attempt to measure public opinion on a great variety of issues: presidential popularity, gun control, abortion, foreign policy, and many more. Education is but a small part of this, and the voucher issue is just a part of education. However well trained these researchers may be in survey methodology, they cannot be expected to have a nuanced understanding of each and every issue. As a result, they may sometimes adopt wording that seems perfectly acceptable, but that misses the mark.

In my view, that is what's happening here. After all, the special-interest wording says something about vouchers that is quite true: a voucher program would indeed provide government subsidies to parents who go private. Moreover, this description is simple and short, properties that researchers value as they economize on survey time. So I'm not surprised that items of the first type have proved popular with general polling organizations. The only problem is that this simple, straightforward approach fails to capture the very purpose of a voucher program.

Given their diverse responsibilities,

Public Opinion Essentially Stable (Figure 4)

Contrary to Phi Delta Kappa's claims, that there has been no "significant decline" in support for vouchers since 1996.



Gallup Poll question: Would you vote for or against a system giving parents government-funded school vouchers to pay for tuition at the public, private, or religious school of their choice?

NBC/ *Wall Street Journal* question: See text, pp. 75-76.

SOURCES: Gallup Poll 1996, 2000; Gallup 2001 experiment; Roper Center 1999, 2000

these researchers can be cut some slack. Their measures are inappropriate, but I suspect they would change them if they gained new perspective on the issue. I can't say the same for PDK's researchers, however. They are responsible, every year, for putting together a survey that deals entirely with education. They know the voucher issue only too well, and they have chosen to measure it in a way that is guaranteed to elicit low numbers.

Has Support for Vouchers Declined?

Now let's return to the issue with which this essay began, the issue of whether there has been a marked drop-off in voucher support over the past few years. PDK claims to have discovered such a downturn. Here is an excerpt from its 2001 press release:

> *It is clear that the decade of the '90s saw support for the use of public funds for parents and students to use in attending private and church-related schools increase, peak, and then begin what has become a significant decline. Support in this area was at 24 percent in 1993, climbed to 44 percent in 1997 and 1998, and has since dropped to the current level of 34 percent.*

The figures, of course, are taken from PDK's "at public expense" item, which is a terrible measure to begin with and consistently gives low scores. Even so, this item has been asked in identical form over time, and the fact that the findings seem to fit a pattern is at least interesting. The question is, does this pattern stand up to scrutiny?

The "significant decline" in voucher support—loudly announced by Phi Delta Kappa and reported by the media as fact—is an artificial phenomenon of Phi Delta Kappa's own making.

The first point to keep in mind is that survey results are likely to fluctuate from year to year by chance alone. Annual shifts of a few percentage points in either direction may be quite meaningless, and analysts have to resist the temptation to overinterpret. Especially in a short time series, random fluctuations can sometimes look like patterns or significant events, when in fact nothing about public opinion has changed.

Based on other surveys, there does seem to have been a real (nonrandom) increase in support from the early 1990s to the mid-1990s, as PDK claims. But beyond the mid-1990s, the evidence does not suggest any pattern at all. In particular, it does not suggest that vouchers have gone into "significant decline" in recent years.

Gallup's own surveys are telling. They are well worded and so should give good measures of how people respond to the basic purposes of a voucher program. Yet there is no indication, by these measures, that support has budged much over the past five years. In 1996 Gallup tested support for vouchers on two separate occasions, and each time came up with support scores of 59 percent. In 2000 it surveyed opinion using the identical question, resulting in support of 56 percent. And in 2001—the experiment—it used a virtually identical question and found support to be 62 percent (see Figure 4). All this is consistent (given normal sampling error) with the notion that public opinion has not changed—and it is surely inconsistent with PDK's claim that vouchers have gone into "significant decline" in the past few years.

Consider another example. In two recent NBC/*Wall Street Journal* polls, respondents were asked to choose between the following positions. "Position A: Government should give parents more educational choices by providing taxpayer-funded vouchers to help pay for private or religious schools. Position B: Government funding should be limited to children who attend public schools." In 1999 the results were 47 percent in favor of vouchers, 47 percent against. But in 2000—at a time when vouchers were supposedly dropping like a stone in popularity—the results were 56 percent in favor and 38 percent against (see Figure 4).

I could provide more examples of vouchers' increasing in popularity over the past few years. I could also find examples of vouchers' declining somewhat in popularity. All of this, however, is quite normal and precisely what we ought to expect if the underlying reality hasn't changed much. Due to sampling error alone, some numbers will go up and others will go down. But the fluctuations probably don't mean much of anything.

PDK Changes the Survey

This is true of voucher surveys in general. The recent plunge in PDK's own measures of voucher support, however, cannot be chalked up to sampling error alone. There is a concrete reason, I believe, why PDK's results have so dramatically gone south. The reason is simply this: PDK has recently *changed* its survey.

Before 2000, PDK followed a format in which respondents were asked to give the public schools a grade from A to F and then were presented with the two voucher items. In 2000, however, PDK altered the survey in a way that any competent researcher would expect to be consequential. Respondents were asked the same A to F grading questions, but then—immediately before the key voucher items—they were asked five additional questions that surely had important framing effects.

The first two essentially set up a dichotomy between vouchers and the public schools. The second of them asks, "Which one of these two plans would you prefer—improving and strengthening the existing public schools, or providing vouchers for parents to use in selecting and paying for private and/or church-related schools?" PDK is thus clearly suggesting to respondents that people who support public education—as most Americans do—cannot at the same time support vouchers. From a framing standpoint, this is a killer. It is also factually incorrect. Most activists in the voucher movement *are* dedicated to improving the public schools, and they see vouchers as a powerful means of effecting improvement through greater choice and competition.

The next three new items are also problematic. Two focus attention on a long list of public-school ideals. The third contrasts parental choice with other "possibilities"—like rigorous academic standards and competent teachers—again giving the impression that they are alternatives to vouchers rather than (as is in fact the case) entirely complementary.

There is little doubt, in my view, that the introduction of these five new items just prior to the usual voucher items produced a more negative framing of the voucher issue and encouraged the lower support scores that became PDK's findings in 2000. The same sort of thing happened in 2001, except in that year PDK's researchers changed the survey again. This time, they eliminated three of the five lead-in items, and included just the two killer items: the ones that portray vouchers as antithetical to public edu-

Molding Public Opinion (Table 1)

It's no wonder that the public's support for vouchers was so low on Phi Delta Kappa's most recent opinion survey. The survey asked a series of questions that framed education reform as a false trade-off between vouchers and fixing the public school system.

Voucher Questions from the 33rd Annual Phi Delta Kappa/Gallup Poll of the Public's Attitudes Toward the Public Schools, 2001

Framing Questions

Survey Question #5: In order to improve public education in America, some people think the focus should be on reforming the existing public school system. Others believe the focus should be on finding an alternative to the existing public school system. Which approach do you think is preferable–reforming the existing public school system or finding an alternative to the existing public school system?

Survey Question #6: Which one of these two plans would you prefer–improving and strengthening the existing public schools or providing vouchers for parents to use in selecting and paying for private and/or church-related schools?

Biased Question

Survey Question #7: Do you favor or oppose allowing students and parents to choose a private school to attend at public expense?

Well-worded Question

Survey Question #8: A proposal has been made that would allow parents to send their school-age children to any public, private, or church-related school they choose. For those parents choosing nonpublic schools, the government would pay all or part of the tuition. Would you favor or oppose this proposal in your state?

cation (see Table 1). Whether this lead-in is more or less negative in its framing than the 2000 lead-in is unclear. I suspect it is more negative, because it is so simple and forceful and avoids all the distractions of the other three items. In either event, it is surely more negative than the original framing from 1999 and before, and there can be no surprise that it again led to lower support scores.

The most reasonable conclusion, therefore, is that the "significant decline" in voucher support—loudly proclaimed by PDK and reported by the media as fact—is an artificial phenomenon of PDK's own making. The important changes didn't occur in public opinion. They occurred in the design of PDK's survey—a factor that, needless to say, is under the conscious control of PDK's own researchers.

Conclusion

I do not want to believe that PDK is using its survey to further its own political agenda. But what is the alternative? That PDK's researchers have simply been careless in their design decisions? That these decisions have by sheer accident led to lower support scores for vouchers? That the most biased of these scores have unwittingly been urged on the media as good-faith evidence of American public opinion? This is a lot to swallow. It is much more reasonable to believe that PDK's researchers are competent at their jobs and that they have not been making one mistake after another.

Whatever the explanation, one conclusion is sure. With the public relations campaign against vouchers in full swing, it is important for people who want the facts about public opinion to look askance at this most official of all education surveys. On the voucher issue, its findings are not to be believed.

Having said this, I don't want to leave the impression that, when the truth of the matter is revealed, Americans turn out to be wild about vouchers. Granted, well-worded survey items do show majority support for the idea. This is very important. But in a population that is poorly informed about public policy, the positions people take on surveys are often soft—and complicated—and need to be evaluated with care.

This is what I have tried to do in my recent book, *Schools, Vouchers, and the American Public*. One of its central themes is that Americans are on both "sides" of the voucher issue at once: they like the public school system *and* they are positively inclined toward vouchers. In their minds (and mine), there is no reason one can't be supportive of both, no reason vouchers can't coexist with and promote a healthy public school system. Problems arise, however, when people are told—as they are during initiative campaigns, through massive media blitzes by the teacher unions—that vouchers will destroy the public schools. Faced with such draconian framing, and responding to the uncertainty that it creates, many people back away from vouchers and embrace the current system. They don't want to lose what they have.

Support for vouchers, then, is a complex matter that cannot be summed up in simple survey numbers. I suspect, moreover, that this complexity may take a still different form over the next year or so, as the tragedy of September 11 generates an upswing in support for government institutions, along with an increasing aversion to change and uncertainty—all of which could show up (temporarily) in the way people respond to survey items about vouchers.

Whatever the future holds, however, the challenge for those of us who are studying this issue—regardless of which side we're on—is to understand what Americans are really thinking. This means insisting on well-designed surveys. And on facts that can be believed.

–Terry M. Moe is a professor of political science at Stanford University and a senior fellow at the Hoover Institution.

Taking Account of Accountability:
Assessing Politics and Policy

June 10–11, 2002

Harvard University
Cambridge, Massachusetts

Keynote Address by
Secretary of Education
Rod Paige

Also presenting:

John Bishop
Anthony Bryk
Ronald Ferguson
Eric Hanushek
Caroline Hoxby
Thomas J. Kane
Tom Loveless
Donald McAdams
Terry Moe

For more information on the conference or for a registration form, please call (617) 495-6954 or go to: www.ksg.harvard.edu/pepg/events.htm

Sponsored by the Taubman Center for State and Local Government and the Program on Education Policy and Governance (PEPG), John F. Kennedy School of Government, Harvard University.

www.ksg.harvard.edu/pepg/

“*The Independent Review* is excellent.”

— **GARY S. BECKER**, Nobel Laureate in Economics

Transcending the all-too-common superficiality of public policy research and debate, ***The INDEPENDENT REVIEW*** is the widely acclaimed quarterly, peer-reviewed journal devoted to excellence in the study of political economy and the critical analysis of government policy. Edited by **Robert Higgs**, ***The INDEPENDENT REVIEW*** is superbly written, provocative, and based on solid peer-reviewed scholarship.

Probing the most difficult and pressing of social and economic questions, ***The INDEPENDENT REVIEW*** boldly challenges the politicization and bureaucratization of our world, featuring in-depth examinations of current policy questions by many of the world's outstanding scholars and policy experts. Undaunted and uncompromising, this is *the* journal that is pioneering future debate!

*“**The Independent Review** is of great interest.”*
— C. VANN WOODWARD, Pulitzer Prize-Winner, Yale University

*“I really enjoy reading **The Independent Review**.”*
— JOYCE APPLEBY, Professor of History, UCLA

*“**The Independent Review** is distinctive in badly needed ways.”*
— LELAND YEAGER, Professor of Economics, Auburn University

“The most exciting new intellectual journal in many years.”
— WILLIAM NISKANEN, Chairman, Cato Institute

“A great team, economic logic and tough analysis.”
— BRUCE YANDLE, Former Executive Director, Fed.Trade Comm.

*“**The Independent Review** is excellent, diverse, and wide-ranging.”*
— IRVING LOUIS HOROWITZ, Prof. of Sociology, Rutgers University

In Recent Issues:

Medicare's Progeny: The 1996 Health Care Legislation
— CHARLOTTE TWIGHT
Does Modernization Require Westernization?
— DEEPAK LAL
Democracy and War
— RUDY RUMMEL AND TED GALEN CARPENTER
Liberty and Feminism
— RICHARD A. EPSTEIN
A New Democrat? The Economic Performance of the Clinton Admin.
— JOHN W. BURNS AND ANDREW J. TAYLOR
A Free Market in Kidneys: Efficient and Equitable
— WILLIAM BARNETT, MICHAEL SALIBA AND DEBORAH WALKER
Racism, Railroad Unions and Regulation
— DAVID E. BERNSTEIN
The School-Choice Choices
— JOHN D. MERRIFIELD
The Immigration Problem: Then and Now
— RICHARD VEDDER, LOWELL GALLAWAY, AND STEPHEN MOORE
Eco-Industrial Parks: The Case for Private Planning
— PIERRE DESROCHERS

✦ Feature Articles
✦ Book Reviews
✦ Special Features
✦ Figures and Tables
✦ 160 Pages per Issue
✦ March, June, Sept. & Dec.
✦ Annual Index
✦ Plus more . . .

Individual Subscriptions:
$28.95/1 year, $52.95/2 year
Institutional Subscriptions:
$83.95/1 year, $158.95/2 year
International Orders:
Add $28/subscription shipping

Order Toll-Free (U.S.)
800-927-8733
Phone: **510-632-1366**
Fax: **510-568-6040**
review@independent.org
http://www.independent.org

The INDEPENDENT INSTITUTE
100 Swan Way, Dept. A65
Oakland, CA 94621-1428 USA

Look for *The INDEPENDENT REVIEW* on better newsstands and in bookstores!

ELECTRONIC REPRODUCTION PROHIBITED

check the facts

by KATE WALSH

Positive Spin

The evidence for traditional teacher certification, reexamined

A constellation of interest groups, including education schools, powerful foundations like Carnegie and Rockefeller, and a variety of professional associations, has been urging states to adopt stricter licensing and certification requirements for teachers. They not only share a deep-seated belief that certified teachers are better teachers; they also claim that their views are supported by a comprehensive body of research. Academics like Stanford University professor Linda Darling-Hammond and associations like the National Council for the Accreditation of Teacher Education (NCATE) and the National Commission on Teaching & America's Future (NCTAF) routinely refer to the "dozens" or "hundreds" of studies showing that certified teachers outperform their uncertified peers. For instance, from an NCATE newsletter published in 1999: "Over 100 studies show that qualified teachers outperform those with little or no preparation in helping students learn." And from NCTAF's 1997 report *Doing What Matters Most: Investing in Quality Teaching*, authored by Darling-Hammond: "More than 200 studies contradict the long-standing myths that 'anyone can teach' and that 'teachers are born and not made.' "

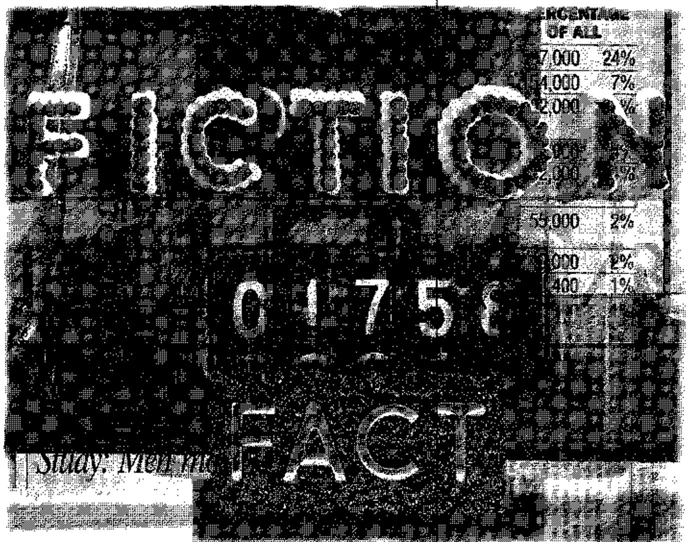


ILLUSTRATION BY JOHN WEBER

Such statements lend a patina of scientific objectivity to the rules and regulations sought by these organizations. After all, who could oppose tighter licensing requirements *if* they were shown to lead to higher student achievement? Moreover, the use of phrases like "hundreds of studies show" makes it seem as though the question is closed, the debate over. Who could doubt findings that have been replicated "hundreds" of times? Checking the veracity of such statements would be a painstaking task, and why doubt them when they're made by respected scholars? Thus there has been no comprehensive effort by scholars in the field to drill down through these layers of evidence to discover what, if anything, they say about the merits of certifying teachers.

One of the primary goals of the Baltimore-based Abell Foundation, where I am a senior policy analyst, is to improve the public schools that serve poor children. For years my colleagues and I have witnessed the frustration of principals who were prevented by the state department of education from hiring promising teaching candidates because they hadn't met the requirements for certification. We wondered if the training provided by schools of education was so valuable that it had to be required, as every state currently does. So we began a search for the evidence linking certification with higher student achievement.

The Maryland Department of Education got us started with 12 documents that the state believes demonstrate the value of certification. We were immediately struck by the paucity of evidence found in these 12 documents and decided to keep digging. Finding many of the studies cited by advocates proved difficult. We often had to track down the author in order to obtain studies that no longer are, or never were, available through a university library. A few unpublished studies ultimately proved impossible to find, but we eventually looked at *every* published study, some going as far back as the 1940s, that supporters of teacher certification cite. We also retrieved many unpublished studies, mostly doctoral dissertations. Even though these dissertations had not undergone the peer-review process that most academic fields consider a prerequisite, we were willing to consider any evidence. In the end, we closely examined the findings of well over 200 studies, literature reviews, meta-analyses, and advocacy pieces.

We focused on research that examines how the various attributes of teachers affect student achievement, counting as legitimate evidence only those studies that used this measure of teachers' effectiveness. Other measures of teachers' effectiveness, such as principals' or other supervisors' ratings of their teachers' performance, are highly unreliable and studies using such measures are often characterized by poor research design.

Certification carries a high price, both financially and in the burdens it imposes on would-be teachers. About one-third of the coursework required of a prospective teacher in college is under the purview of a school of education. Certification requirements also dictate how states will appropriate millions of tax dollars to universities and colleges, and they determine the structure of incentive programs so critical to school districts in times of teacher shortages. Access to the profession is littered with obstacles for some of the most promising candidates: middle-aged professionals who wish to switch careers and talented college graduates who didn't major in education. Even in states that appear to offer a bypass around certification, these alternative programs are structured so that retrofitting—taking education coursework while teaching—is almost always required. To justify making the path to the teaching profession so difficult, certification needs to yield significant benefits in terms of student learning. Anything less, and states ought to question whether education schools should be the predominant path to becoming a teacher, rather than just one of many.

Some of the most oft-cited studies had such serious flaws that no properly trained researcher would take them seriously.

The Evidence

The body of research on the effects of teacher certification is astonishingly poor. Some of the most oft-cited studies had such serious flaws that no properly trained researcher would take them seriously. Early in our project we were amazed by the extent to which research is misquoted, misinterpreted, and misrepresented. Several researchers complained that their work, though widely cited, did not appear ever to have been read, considering the conclusions that were reached.

The examples presented here illustrate the seven main ways in which the supporters of teacher certification have twisted and turned the research findings to suit their policy purposes. The entire analysis of more than 200 studies and citations to the studies listed here can be found in the full study and in the unabridged web version of this essay, at www.educationnext.org.

✦ *Research that seems to support teacher certification is selectively cited, while research that does not is overlooked.*

In the 1997 NCTAF report *Doing What Matters Most,* Darling-Hammond describes the findings of a study (Ferguson 1991) that examined various characteristics of teachers and their relationships to student achievement. The study demonstrated that certain attributes play an important role in raising student achievement. Darling-Hammond cites this study as proof of the paramount importance of a teacher's possessing knowledge about "teaching and learning." As a result, she urges states to invest in more formal teacher education in order to ensure that teachers obtain this knowledge. She claims that Ferguson found that "every additional dollar spent on more high-qualified teachers netted greater increases in student achievement than did less instructionally focused uses of school resources."

Yet Ferguson's study did not show that a teacher's pedagogical training had an effect on student achievement. It was teachers' *verbal ability* that had a very large effect. At no point did Ferguson suggest that his research argues for states investing in more formal teacher education, though he did try to figure out how money could be used to attract teachers with higher SAT scores.

To measure teacher's verbal ability, Ferguson used a literacy test that was administered in 1986 to all public school teachers in Texas. He found a surprisingly large correlation between how well teachers did on this relatively easy test (the pass rate was 97 percent) and their students' achievement on a standardized test. Ferguson distinguishes this literacy test from tests like the National Teacher Exam (NTE), because it did not test pedagogical knowledge, a distinction that Darling-Hammond ignores. He specifically notes that previous research has found "scarce and weak" evidence of a relationship between tests of pedagogical knowledge and student achievement. Darling-Hammond wrongly states that the Texas test was a licensing test (actually it was a recertification test given to already hired teachers) and "that it tested both basic skills and teaching knowledge."

Why does Darling-Hammond feel justified in claiming that the test proves the importance of teaching knowledge? In fact, the Texas test did contain a section entitled "professional knowledge." However, Ferguson (and, indeed, anyone who ever took the test) understood this section as a test of basic vocabulary. In a series of ten multiple-choice questions, teachers were asked to pick the right definition of such phrases as "standardized tests," "classroom management," and "certification." It is certainly incongruous to suggest, as Darling-Hammond does, that knowing the definitions of these basic terms provides proof that prospective teachers benefit from sitting through 30 credit hours of education coursework or that states should invest their resources in even more formal training.

Darling-Hammond cites another study (Greenwald, Hedges, and Laine 1996) in support of her contention that states need to increase their investments in teacher education. She asserts that the study found that "spending on teacher education swamped other variables as the most productive investment for schools."

Yet, after reviewing 60 studies, Greenwald, Hedges, and Laine's firmest and most significant conclusion was that verbal ability had by far the most significant effect on student achievement. In a separate e-mail, one of the authors confirmed this finding: "Teacher ability (which was generally measured as teacher's verbal ability)," Hedges wrote, "seems to show the strongest and most replicable effect on achievement." However, Darling-Hammond, as she so often does, overlooks this central finding. Her reference to spending on teacher education was based on a tangential exercise the authors engaged in, in which they speculated on how school resources could be best invested in order to improve student achievement, including what would happen if more money were invested in teacher education.

But Darling-Hammond never acknowledges a key failing of this exercise. The authors stated that they were unable to come up with particular ways in which school districts could spend money to improve the average verbal ability of their teachers (though other researchers such as Ferguson and Manski have suggested that higher teacher salaries might do so), so they left out possible ways that money might be spent to raise verbal ability. Teacher education was only able to "swamp" all other spending variables because the authors couldn't figure out how to improve the most important teacher attribute.

◆ *Analyses are padded with imprecise measures in order to conceal the lack of evidence in support of certification.*

Darling-Hammond, in *Teaching and Knowledge: Policy Issues Posed by Alternative Certification for Teachers* (1992), asserts that there are "consistently positive relationships between student achievement in science and the teacher's background in both education courses and science courses," and she cites four studies in support of this statement. In fact, not one of the four studies looked at education coursework as an independent variable. Of course, if a researcher only reports that the combination of education coursework and science coursework improves student achievement, it is impossible for the reader to discern if both or only one of the variables were responsible for the effect.

In particular:

Davis (1964) studied the effect of teachers' science coursework and their participation in National Science Foundation summer institutes. It did not study the effect of teachers' coursework in education.

Taylor (1957) found an effect only when education coursework was combined with science coursework. No effect from education coursework was found when the variable was isolated.

Studies find that teachers' verbal ability has by far the most significant effect on student achievement—certainly more than pedagogical training.

Druva and Anderson (1983) found that science coursework improved student achievement significantly, but that education coursework had no effect when the variable was isolated.

Perkes (1967) can be considered a stand-off at best. Students of teachers with more education coursework under their belts scored higher on a test of higher-order thinking, but lower on a science achievement test.

Another example: in *Teacher Quality and Student Achievement: A Review of State Policy Evidence* (1999), Darling-Hammond reviews what the research says about the relationship between student achievement and many different teacher variables, including teacher's general academic ability, intelligence, subject-matter knowledge, pedagogical knowledge, experience, and certification status. Her intention is to systematically prove her central thesis, that "measures of teacher preparation and certification are by far the strongest correlates of student achievement." While conceding that teachers' knowledge of their subject is important, she claims that the findings relating subject-matter knowledge to teacher effectiveness "are not as strong and consistent as one might suppose." She states that five studies have found "no consistent relationship between the subject-matter tests of the National Teachers' Exam and teacher performance as measured by student outcomes or supervisory ratings."

This is misleading. Not one of the five studies found a negative relationship between the NTE subject-matter tests and student outcomes. First it is important to understand the distinction between the Core Battery portion of the NTE and the subject-matter portion of the NTE, both of which have been replaced by the Praxis exams. The Core Battery was a test of teachers' basic skills and knowledge of pedagogy. The subject-matter portion was a test of a teacher's knowledge of the subject area that he or she was going to teach.

Andrews, Blackmon, and Mackey (1980) found a *positive* relationship between teachers' scores on the NTE English and elementary subject-matter tests and supervisors' ratings. The only *negative* relationship the authors found was between teachers' scores on the NTE physical education and special-education tests and supervisors' ratings of their performance.

Ayers and Qualls (1979) examined the relationship between teachers' scores on the NTE and how their students rated their performance. But students' ratings are extremely unreliable; students may rate a teacher who gives them no homework higher than a teacher who challenges them.

Quirk, Witten, and Weinberg (1973), a literature review on the effect of the NTE, found only one study that looked at the relationship between the NTE and student achievement, and it was written in 1947. Though too old to be considered very relevant, it reported a *positive* relationship between the teachers' scores on the NTE and student achievement.

Haney, Madaus, and Kreitzer (1987) is another literature review, recycling the same findings found in Ayers and Qualls and in Quirk, Witten, and Weinberg. (One of the authors, George Madaus, told us that he was not aware of any research showing a negative correlation between the NTE subject-matter test and student achievement.)

Many assertions about the benefits of certification are largely dependent on the evidence provided in work that has not been peer reviewed.

Summers and Wolfe (1977) was the only study of the five to explore the relationship between teachers' performance on NTE subject-matter tests and student achievement. They found a largely *positive* effect. (The negative effect to which Darling-Hammond refers was probably what Summers and Wolfe noted as the "perversely" negative relationship between 6th grade teachers' scores on the NTE Core Battery, a test of pedagogy and basic skills, and their students' achievement.)

✦ *Researchers focus on variables that are poor measures of the qualities they are interested in, sometimes ignoring variables that are better measures.*

In a 1985 literature review that is still extensively cited, Evertson, Hawley, and Zlotnik argued that the concern with whether teachers know their subject matter is overblown. They cited three studies that supposedly found that there is "no or [a] negative relationship between teacher knowledge (as measured by GPA and standardized tests) and student achievement."

One of these three studies is a 1977 dissertation by Eisenberg. It's true that Eisenberg failed to find a correlation between teachers' GPAs and their students' achievement, but GPA is considered a fairly crude measure of a teacher's subject-matter knowledge; teachers presumably take courses in college other than those in their subject area, courses that might drag their GPAs down. Evertson, Hawley, and Zlotnik neglect to report that Eisenberg found a significantly positive effect from a better measure of teachers' knowledge of mathematics: their knowledge of algebraic concepts and postgraduate coursework in calculus. The other two studies that Evertson, Hawley, and Zlotnik cited—both unpublished—were similarly misreported.

Instead of using standardized measures of student achievement, advocates sometimes design their own assessments in order to prove certification's value. Ashton and Crocker (1983) cited Copley's 1975 dissertation as evidence that beginning teachers are better prepared by virtue of having taken coursework in pedagogy and educational methods. The assessment Copley used was a survey designed by doctoral students in education. It asked principals to rate their new teachers on factors that are unrelated to student achievement, such as how often they participated in professional meetings, if they have a good disposition, and if they have a sense of humor.

Sometimes, when a study doesn't use measures of student achievement, advocates simply say it does. Denton and Lacina's 1984 study is cited repeatedly by advocates of certification because it was supposed to have found a positive relationship between formal teacher preparation and student achievement (Evertson, Hawley, and Zlotnik 1985; Darling-Hammond 1999). Yet Denton and Lacina never even looked at student achievement; their study measured only the morale of student teachers and the ratings of student teachers by their supervisors.

✦ *Research is cited that is old and often irretrievable.*

The fact that research is relatively old does not automatically make it irrelevant. Yet older studies should be regarded skeptically. There are a number of reasons: 1) student achievement probably wasn't used as the measure of teacher effectiveness; 2) before the advent of the modern computer, in the mid-1960s, some of the more sophisticated analyses were not feasible; 3) the structure and makeup of schools change, making the findings less applicable to the current situation; 4) most important, older studies may not control for critical variables, such as students' backgrounds or past achievement.

In *Reforming Teacher Preparation and Licensing: Debating the Evidence* (2000), Darling-Hammond cites numerous studies to support the statement: "Knowledge about teaching and learning shows even stronger relationships to teaching effectiveness than subject-matter knowledge." However, only four of the studies were published after 1980; the remaining 12 were published between 1950 and 1979.

◆ *Research that has not been subjected to peer review is treated without skepticism, and there is a particularly heavy reliance on unpublished dissertations.*

The process of peer review, having researchers' fellow experts review a study before it is deemed worthy of publication, is practiced in all fields of serious scientific study. Yet many assertions about the benefits of teacher certification are largely, if not exclusively, dependent on the evidence provided in unpublished dissertations, papers delivered at conferences but never published, or articles published in the many education journals, like *Phi Delta Kappan*, that are not "refereed" (they don't have a system of blind peer review).

For instance, Ashton and Crocker (1987) cite numerous studies on teacher preparation to support their conclusion that coursework in education makes teachers more effective than coursework in their subject matter does. They claim that 9 of the 14 studies they found showed that subject-matter coursework made no difference. It took a careful read of the footnotes to discover that all but two of these nine studies were dissertations, unpublished and unavailable to scrutinize because they ranged from 25 to 45 years old. The two published studies were equally problematic. Rothman, Welch, and Walberg (1969) studied only 35 teachers who were accepted into an elite project developed by the Harvard Physics Project. Generalizing the findings to the greater population of teachers is inappropriate. The other study (Perkes 1967) produced mixed results: students whose teachers took more subject-matter coursework reported higher scores on an achievement test, but lower scores on the STEP, a test of higher-order thinking. Another example: Druva and Anderson (1983) reported a largely positive link between education coursework and "successful teaching," but 54 of the 65 studies reviewed were dissertations or unpublished articles, all written between 1966 and 1975.

The difficulty of tracking down some of these studies is worth noting. We tried to find one frequently cited unpublished paper, delivered at a 1990 American Educational Research Association (AERA) conference in Boston, which was written by Gomez and Grobe. It was not available from the archives of the AERA and could not be located through the services of a university library. Even the authors no longer had a copy of the paper, and none of the researchers who cited the study was able or willing to produce the report after numerous requests. The authors' recall of their findings did not correspond to Darling-Hammond's presentation of their findings.

◆ *Studies that support teacher certification routinely violate basic principles of sound statistical analysis that other academic disciplines take for granted; methodological errors go unchallenged.*

First, studies often do not control for variables that are critical to understanding student performance. For example, Darling-Hammond, in NCTAF's 1997 report *Doing What Matters Most*, asserts, "Students will achieve at higher levels and are less likely to drop out when taught by certified teachers." She supports this claim with three studies, none of which controlled for students' family backgrounds. Studies of teachers' influence on student achievement need to include controls for students' socioeconomic background, as this variable appears more important than any other variable in determining student achievement. (When citing these studies a second time in the 1999 report *Teacher Quality and Student Achievement: A Review of State Policy Evidence*, Darling-Hammond acknowledged this fact. It would have been more appropriate not to mention these studies at all.)

Second, conclusions are sometimes based on sample groups that do not mirror the general population. The work of Edward Begle, a respected mathematician, is one of eight studies cited by Evertson, Hawley, and Zlotnik (1985) as showing that little evidence exists to support increasing teachers' knowledge of their subjects beyond what is typically required for certification. It's true that Begle (1972) found that the number of mathematics courses taken by a teacher stopped making much of a difference at a certain point. But he points out a critical limitation of these findings that others disregard, including Evertson, Hawley, and Zlotnik: the teachers in the study were an elite group of science teachers who had been accepted by the National Science Foundation Summer Institute. They also felt comfortable enough with their knowledge of mathematics to volunteer to take the math test used for the study. It should come as no surprise that variation in the number of math courses taken by teachers in this group would not lead to different outcomes in student achievement. Differences in the number of math courses taken by teachers in the larger population of science teachers could still lead to variation in student achievement.

Third, conclusions are sometimes

The certification process is incapable of providing any insight into an individual's ability, intellectual curiosity, creativity, affinity for children, and instructional skills.

based on samples that are simply too small to produce results that are reliable or that can be generalized to the larger population. The findings from a study that included only a few teachers cannot yield reliable findings that apply to a larger population. Here is a sampling of studies on certification and the number of teachers that were involved: *Bullough, Knowles, and Crow*, 3 teachers; *Davis*, 29 teachers; *Eisenberg*, 28 teachers; *Grossman*, 3 teachers; *Hall*, 38 teachers; *Hice*, 40 teachers; *Howe*, 51 teachers; *Hurst*, 55 teachers; *Lins*, 27 teachers; *Perkes*, 32 teachers; *McNeil*, 38 teachers; *Rothman, Welch, and Walberg*, 35 teachers; *Thoman*, 29 teachers.

Fourth, some studies suffer from serious statistical errors known as aggregation bias that can lead to misinterpretations. For instance, in 1999, Darling-Hammond published a widely publicized study of the relationship between student performance on the 4th and 8th grade National Assessment of Educational Progress (NAEP) and the certification status of each state's teachers. She found that states with higher student achievement also employed a greater percentage of certified teachers. The problem is the inability to match students with their teachers when using data that are aggregated at the state level. States differ in ways that make drawing conclusions about causation quite troublesome. For instance, states with higher percentages of certified teachers may also have strong accountability systems that focus their schools on student achievement. Or the families in states with higher percentages of certified teachers may place more emphasis on education, thereby raising student achievement. In short, many variables besides teacher certification might explain why scores were higher in some states than in others, but Darling-Hammond was able to control for very few of them.

In the study, Darling-Hammond acknowledges the likely distortions in her analysis: "Aggregating data to the state level produces different results than one would find if one looked at similar kinds of data at the individual student, teacher, school, or district level." However, even the concession that the findings probably aren't accurate doesn't stop her from maintaining that the data are still useful "for the purposes of assessing broad policy influences at the state level." In effect she is stating that the data, accurate or not, should be sufficient for making public policy.

Independent Confirmation

The Abell Foundation study is not the first to uncover the lack of rigorous and legitimate evidence on teacher certification. Charged by the U.S. Department of Education with combing the existing research on teacher preparation and subjecting it to scientific standards used in other fields, scholars with the Center for the Study of Teaching and Policy at the University of Washington (Wilson, Floden, and Ferrini-Mundy 2001) eliminated all but 57 studies written in the past 20 years. However, most of these 57 studies were "interpretive" case studies involving only a few teachers. The actual number of longitudinal or quasi-longitudinal studies that controlled for students' backgrounds and used student achievement as the measure of whether teacher preparation made a difference was far fewer. Under the scholars' own standard, only six studies containing any evidence that teacher certification is effective were left standing, a fact that was omitted in Wilson, Floden, and Ferrini-Mundy's text.

Supporters of certification attribute the low student achievement in the nation's poorest school districts at least partially to the high number of uncertified teachers working in these districts. However, this view remains unsupported by sound research. Moreover, the even lower number of certified teachers employed by the nation's inner-city parochial schools and elite private schools would appear to discount this theory. If certification were crucial, certainly these schools would privilege certified teachers in order to satisfy their tuition-paying families. The theory that teacher certification leads to high-quality teaching is based more on what we think ought to be true (shouldn't coursework in pedagogy and educational methods create better teachers? Shouldn't teachers have to go through education school, just as lawyers go to law school and doctors go to medical school?) than on controlled experimentation. It is a leap of faith taken without the benefit of supporting evidence. The evidence, it turns out, is astonishingly deficient.

Reduced to their essence, the existing routes to teacher certification consist of little more than counting course titles on the college transcripts of prospective teachers. The process is incapable of providing any insight into an individual's ability, intellectual curiosity, creativity, affinity for children, and instructional skills. Acting as a very crude proxy for teacher quality, the process is incapable of identifying significant, justifiable reasons for denying uncertified candidates access to the profession. A highly able candidate who does not take a required course is no more allowed to teach than the candidate who is poorly educated and unable to pass the teacher's examination.

Determining who is qualified to teach is a task fraught with ambiguity and nuance, far more difficult than the mechanical process of counting a teacher's coursework suggests. Given the faulty principles on which certification is based, it is not surprising that its value cannot be proven. Regulatory policy cannot supplant the need for human judgment.

–Kate Walsh is a senior policy analyst at the Baltimore-based Abell Foundation.

Democracy in Education

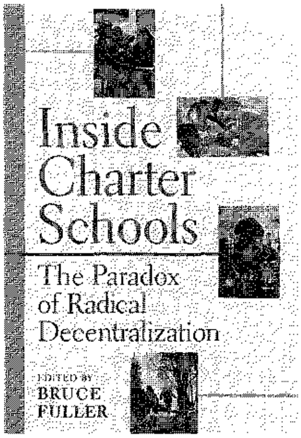
America's latest voluntary organization: the charter school

Inside Charter Schools: The Paradox of Radical Decentralization

Edited by Bruce Fuller

Harvard University Press, 2000, $31.50; 288 pages.

Reviewed by Patrick J. Wolf



The soaring popularity of charter schools among parents, education reformers, and politicians still hasn't convinced Bruce Fuller of their worth. Fuller, a professor of education and public policy at the University of California at Berkeley, claims that the charter schooling movement is paradoxical in that it undermines the very values—of equity and community—that it seeks to promote.

Citing French philosopher Jean-Jacques Rousseau and sociologist Emile Durkheim, Fuller holds that modern democratic societies depend on government-generated and government-enforced social cohesion. In this vein, modern governments need to resolve disputes over how children ought to be educated and then use their authority to impose a consistent standard on all. Fuller sees charter schooling as a "radical decentralization" of authority whereby decisions regarding the education of children and the allocation of public resources are improperly entrusted to groups of parents, advocates, and charter school leaders.

These groups, Fuller claims, are concerned only with their particularistic self-interests and lack a proper regard for the "common good." The result, Fuller suspects, is the creation of public schools that are public in name only. They act as gated communities for the privileged, contrasted with the traditional common public schools where children of diverse races, social classes, and abilities are mixed and forged into responsible democratic citizens. In Fuller's words, "The charter movement—now sanctified as the centrists' rendition of school choice—may contribute to the dismantling of the modern state's political foundations."

An earlier analyst of American society, Alexis de Tocqueville, considered such voluntary community organizations the mainstay of democracy in America.

Fuller repeatedly accuses charter school advocates of being "tribal" and particularistic in their interests. Yet, in describing the left-right coalition that has developed in support of charters, he admits that it is unified behind a broad view of the public interest. In fact, the charter school "cabal" that Fuller considers so worrisome is precisely the sort of broad-based and cross-cutting coalition that the writers of *The Federalist* believed would be the proper engine of policy development and change in the United States. Unlike the Founders, Fuller wants America to be one big unitary tribe—all unum and no pluribus—with a strong central government as "chief." Fuller views parent and grassroots advocacy groups as instruments of political fragmentation and social discord. An earlier analyst of American society, Alexis de Tocqueville, considered such voluntary community organizations the mainstay of democracy in America.

Beyond his philosophical objections, Fuller even doubts whether the supposedly privileged students who enroll in charter schools benefit academically from the experience. He is convinced that the burgeoning charter school movement is a false panacea that merely distracts decisionmakers from the kinds of social changes—equalizing education funding and eliminating poverty—that are the keys to better learning.

Fuller purports to be "concerned about the lack of dispassionate evidence" regarding charter schools, and he offers this collection of case studies (written by his graduate students, fellow UC professors, and an education writer for the *New York Times*) as a partial remedy. Yet Fuller's own criticisms of school choice involve more passion than logic. The research presented here adds little to the body of credible evidence on the performance of charter schools relative to traditional public schools.

Fuller and his collaborators question the evidence that charter schooling is superior to traditional public schooling. Granted, as yet there is no rock-solid scientific evidence that students benefit from a switch from traditional to charter public schools. The bulk of the evidence supporting charter schools is found in surveys of parental satisfaction and in the undeniable fact that an increasing number of parents are placing their names on the waiting lists of oversubscribed charter schools. They certainly seem to believe that charters offer a better alternative. Regardless, this empirical void would seem to have given Fuller the perfect opportunity to conduct a more systematic test of charter school effectiveness. It's unfortunate that he didn't take it.

Instead, Fuller's research team merely visited several individual charter schools and described their histories, characteristics, and the opinions of their principals, teachers, parents, and students—warts and all. Such case studies are a valuable tool for generating research hypotheses or exploring how policy interventions are being shaped in the trenches. The researchers documented plenty of disappointing shortcomings in the charter schools they studied. Based on the case studies presented here, we might responsibly conclude that some, perhaps many, charter schools are imperfect. However, we cannot conclude, as Fuller does, that charter schools on the whole are undesirable alternatives to traditional public schools. We should not draw any firm conclusions from the seven cases described in this study because they are too few in number to allow generalizing to all of the nation's 2,372 charter schools, a problem that Fuller himself acknowledges.

In addition, the book provides no full-fledged case studies of traditional public schools whose strengths and weaknesses can be compared with those of the seven charter schools presented here. I am aware of no school choice advocate who claims that schools of choice are flawless—only that they are, on average, better than the alternative. Since Fuller and his colleagues give us no information on the schools that these students left behind in order to attend charter schools, their study offers no empirical foundation from which to draw general conclusions of any kind.

Fuller insists that school choice schemes, including charter schools, threaten America's "social organization" and diminish the civic values of students. But he marshals no systematic evidence in support of this assertion. And it is more likely, in my view, that students who are placed in schools of choice, rather than compelled by the government to attend a particular school, will turn out to be less authoritarian and more politically tolerant and community-spirited than their public school peers. Why would we expect, as Fuller does, that an education system that coercively assigns students to schools in ways that benefit primarily upper-income families would necessarily promote an appreciation for liberty, democracy, and equality? Why would we expect that a system that confines the residents of racially segregated inner-city neighborhoods to racially segregated government schools would somehow promote an appreciation for diversity and pluralism in America?

We should not draw any firm conclusions from the seven cases described in this study because they are too few in number to allow generalizing to all of the nation's 2,372 charter schools.

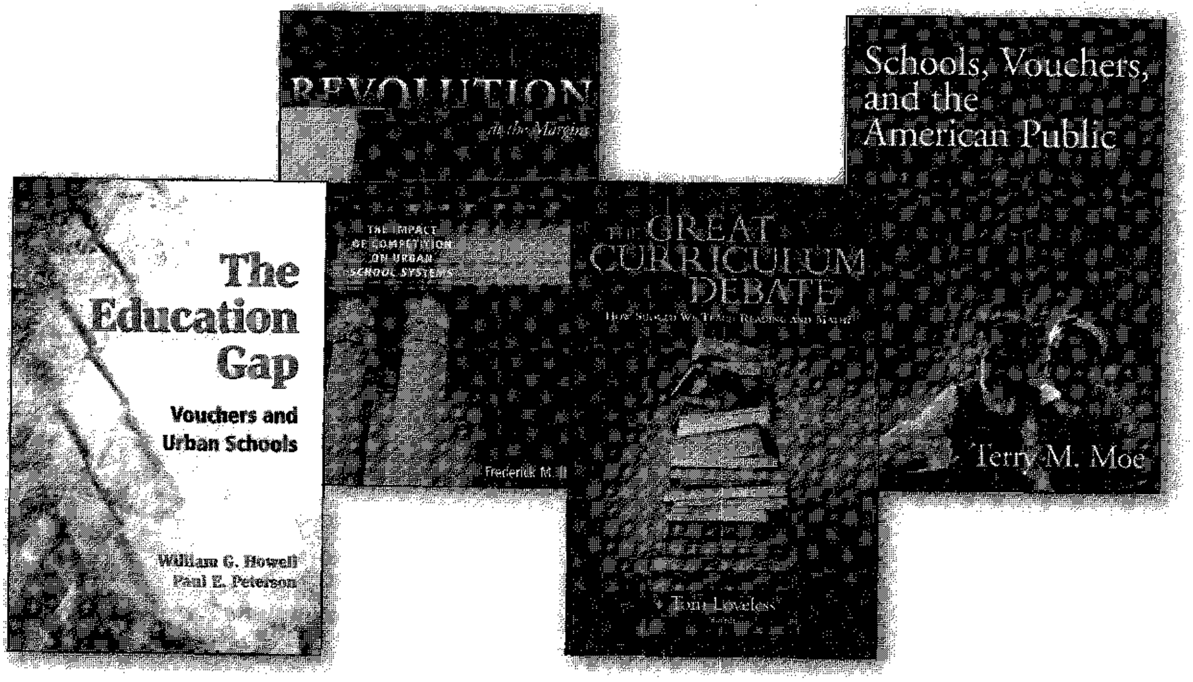
Fuller asks, "Is this creation of small, even tribal, public squares bounded by ethnicity, social class, or religion a viable way to enrich communities? Or is it part of a political agenda bent on eroding the larger public square, the foundations of a common good to which the modern state has been committed for more than two centuries?" Fuller answers his rhetorical questions with the claim that, under school choice regimes, "The state sanctions the pursuit not of the broad common good but of private interests."

However, a basic understanding of American history undermines Fuller's perceptions of the altruistic common school tradition and the threat to the "common good" posed by school choice. There was no public school system 200 years ago. To the extent that formal education existed in America, it was provided by parents or private schools. Moreover, the public school system that emerged in Massachusetts in the 1840s was arguably as particularistic as the major private or charter school systems of today. Horace Mann's famous public schools promoted a distinctly Unitarian brand of Protestantism and at times were virulently anti-Catholic—a reality that Fuller only alludes to in order to impugn the motives of Catholics who currently support school choice.

The work of Charles Glenn of Boston College has shown that the American public school system never truly embraced a secularized "common school" conception of the public good. Protestant prayers were recited in public schools until the 1960s. Students were assigned to racially segregated public schools in the South until the late 1950s. A majority of African-American public school students are still educated in schools that are de facto racially segregated. The truth is that grassroots organizations like a community-developed charter school are far more traditional in America than the bureaucratic structures of the modern public school system.

–Patrick J. Wolf is an assistant professor of public policy at Georgetown University and a member of the National Working Commission on Choice in K–12 Education.

BROOKINGS INSTITUTION PRESS
AHEAD OF THE CURVE


REVOLUTION
At the Margins
THE IMPACT OF COMPETITION ON URBAN SCHOOL SYSTEMS
Frederick M. Hess
Schools, Vouchers, and the American Public
Terry M. Moe
The Education Gap
Vouchers and Urban Schools
William G. Howell
Paul E. Peterson
THE GREAT CURRICULUM DEBATE
How Should We Teach Reading and Math?
Tom Loveless

The Great Curriculum Debate
How Should We Teach Reading and Math?
Tom Loveless, ed.

Charters, Vouchers, and Public Education
Paul E. Peterson and David E. Campbell, eds.

Evidence Matters
Randomized Trials in Education Research
Frederick Mosteller and Robert Boruch, eds.

The Education Gap
Vouchers and Urban Schools
William G. Howell and Paul E. Peterson

The 2001 Brown Center Report on American Education
How Well Are American Students Learning?
Tom Loveless, Director

Revolution at the Margins
The Impact of Competition on Urban School Systems
Frederick M. Hess

Brookings Papers on Education Policy: 2002
Diane Ravitch, ed.

When Schools Compete
A Cautionary Tale
Edward B. Fiske and Helen Ladd

Spinning Wheels
The Politics of Urban School Reform
Frederick M. Hess

Fixing Urban Schools
Paul T. Hill and Mary Beth Celio

It Takes a City
Getting Serious about Urban School Reform
Paul T. Hill, Christine Capmbell & James Harvey

The Black-White Test Score Gap
Christopher Jencks and Meredith Phillips, eds.

All Together Now
Creating Middle Class Schools through Public School Choice
Richard D. Kahlenberg
A Century Foundation Book

Conflicting Missions?
Teachers Unions and Educational Reform
Tom Loveless, ed.

The Tracking Wars
State Reform Meets School Policy
Tom Loveless

NOW IN PAPERBACK

Choosing Equality
School Choice, The Constitution, and Civil Society
Joseph P. Viteritti

Schools, Vouchers, and the American Public
Terry M. Moe

Order from
Brookings Institution Press
Phone: 800-275-1447
202-797-6258
Fax: 202-797-2960
Or order online at
www.brookings.edu

education matters to me

by JEFFREY JONES

Low Expectations

Straight A's, yet not ready for college

My high school was certified as "college preparatory." I was able to take introductory calculus, advanced chemistry and biology, and even several English literature courses for college credit. I graduated as valedictorian of my decent-sized class, with just over a 3.9 GPA.

However, as my father often observed with great frustration, I rarely spent my evenings doing homework; I was able to finish my work during class or in study halls. What bothered him was that I wasn't being challenged by my high school. So he thought that I should go beyond the assigned workload. I considered the request ridiculous—what with my emerging NBA career just getting under way and all. Besides, it wasn't as if I were a mediocre student; I was the best in the entire class.

PHOTOGRAPH COURTESY OF JEFFREY JONES

After graduation, I decided to attend the University of Wisconsin–Madison, the flagship of the UW system. I was a little cocky going in; I figured that if anyone in my entire high school was prepared for college, I was. My experience, however, suggests that none of my high-school classmates was prepared for college. My first few years at Madison were disastrous. Whereas in high school I had taken introductory calculus, I was forced into remedial math at Madison (not even introductory algebra or geometry). I fared no better in the physical sciences. Three of my high-school friends eventually dropped out, never to graduate from college.

After floundering initially, I resolved to do whatever was necessary to do well. I spent many hours in professors' offices, tutorials, and reading groups. I eventually recovered, with quite excellent grades in my final two years. But one cannot entirely recover—GPA-wise—from such a poor start.

My undergraduate experience was both humbling and belittling. Even today—a Ph.D. in philosophy under my belt, a J.D. in the works—my academic confidence remains fragile. My high school created a false sense of aptitude and competence and failed to provide me with the foundation on which all of my secondary learning was to be built.

My high school created a false sense of aptitude and competence.

It is often thought that if children simply work hard and are supported by their parents, they can overcome substandard schooling. It's true that many students do excel despite their woeful schools. But more prevalent are cases like mine: students for whom the rigors of college life were a rude awakening, to whom their schools did an enormous disservice by expecting less than they were capable of.

Personal initiative and parent support are indispensable pieces of the educational jigsaw. It is a mistake, however, to treat them as sufficient. It is equally important that schools enforce standards of educational excellence at every grade level. In the absence of an unflinching, daily reaffirmation of standards of educational excellence, the structural disadvantages common to many public schools—the poorly paid and underqualified teachers, the crumbling facilities—inevitably begin to encroach on, and then to corrupt, the learning environment.

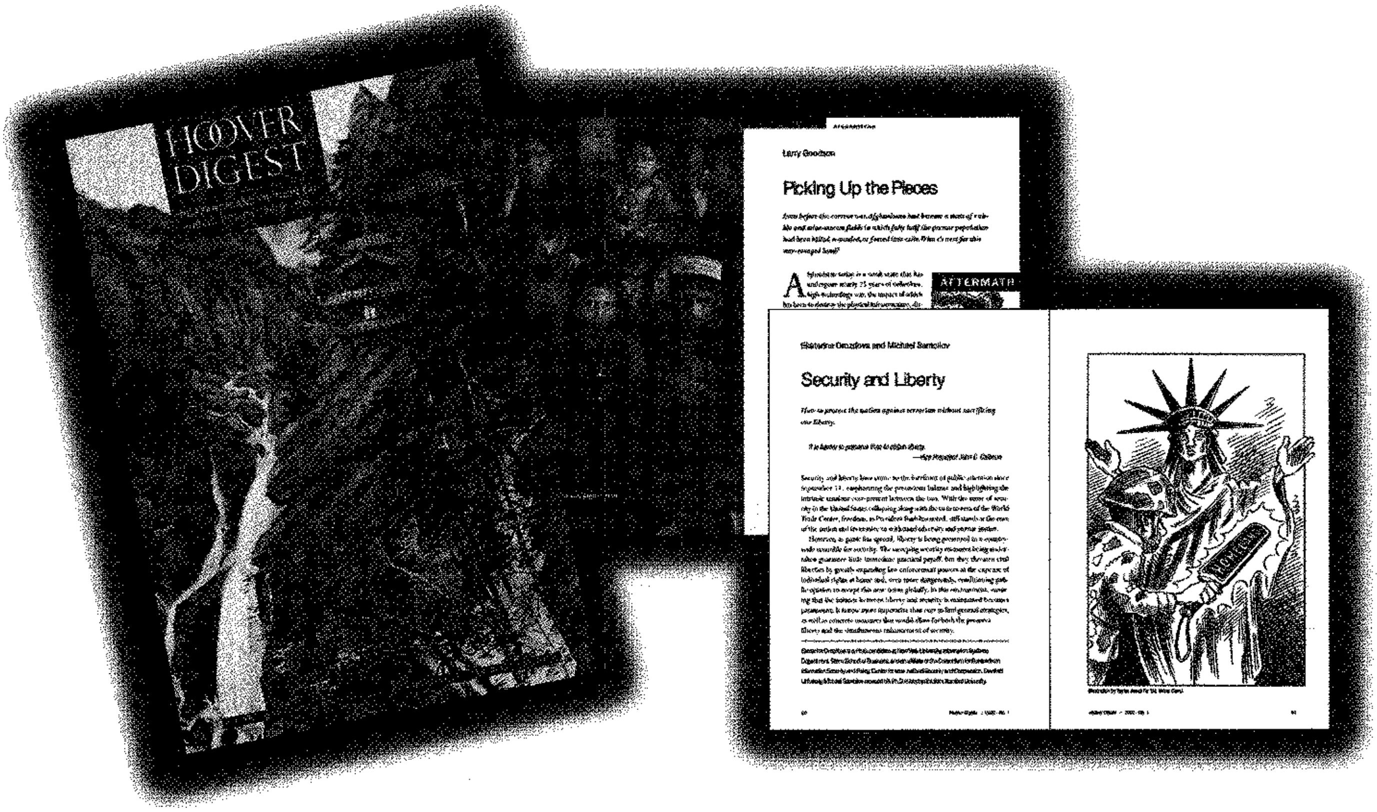
My educational biography speaks to no particular proposal for education reform. However, there is one reform that my story might seem to support, but which may do more harm than good. One way of enforcing standards of excellence is through standardized testing, linked to high stakes—such as holding students back until they pass the tests or delaying their graduation.

This would be fine if their schools were receiving the funds necessary to hire good teachers and to create an inspiring environment for learning. Otherwise children are just being held hostage to their schools' failings. Unless we transform urban schools, it may be better simply to let students move on and find their way. I did.

–Jeffrey Jones is an assistant professor of philosophy at the University of Massachusetts–Boston. He is currently on leave attending the University of Michigan Law School.

HOOVER
DIGEST

Informative. Provocative. Insightful.


HOOVER DIGEST
AFTERMATH
Picking Up the Pieces
Security and Liberty

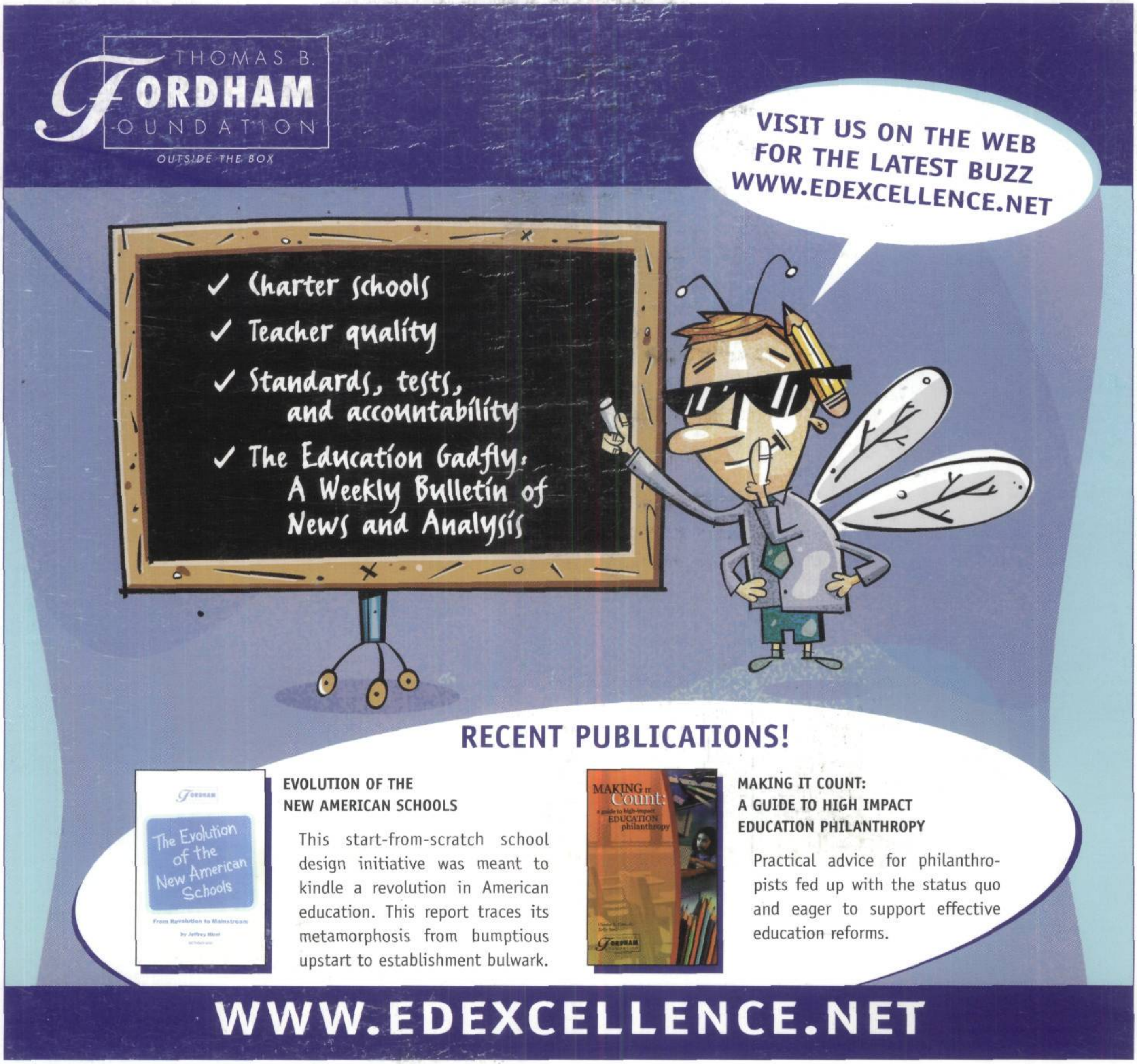
Turn to the quarterly *Hoover Digest* for lively and compelling writing on politics, economics, history, and culture from the Hoover Institution, one of the nation's preeminent think tanks.

A Free Issue with No Obligation! Call now and receive a free copy of the latest issue. If you like what you see, you can subscribe at the special introductory rate of $20.

TO RECEIVE YOUR FREE ISSUE, CALL 800-935-2882.

Hoover Institution, Stanford University, Stanford, California 94305-6010 fax: 650-723-8626
www.hooverdigest.org

ELECTRONIC REPRODUCTION PROHIBITED



THOMAS B. FORDHAM FOUNDATION
OUTSIDE THE BOX

VISIT US ON THE WEB FOR THE LATEST BUZZ WWW.EDEXCELLENCE.NET

- ✓ Charter schools
- ✓ Teacher quality
- ✓ Standards, tests, and accountability
- ✓ The Education Gadfly: A Weekly Bulletin of News and Analysis

RECENT PUBLICATIONS!

EVOLUTION OF THE NEW AMERICAN SCHOOLS

This start-from-scratch school design initiative was meant to kindle a revolution in American education. This report traces its metamorphosis from bumptious upstart to establishment bulwark.

MAKING IT COUNT: A GUIDE TO HIGH IMPACT EDUCATION PHILANTHROPY

Practical advice for philanthropists fed up with the status quo and eager to support effective education reforms.

WWW.EDEXCELLENCE.NET

EDUCATION next
A JOURNAL OF OPINION AND RESEARCH

226 Littauer North Yard
1875 Cambridge Street
Cambridge, MA 02138-3001
Tel. (617) 495-2491, Fax (617) 496-1507
Website www.educationnext.org

To subscribe or to submit change of address information, visit us at www.educationnext.org; fax us at (650) 723-8626; email us at hooverpress@hoover.stanford.edu; or write to *Education Next* Subscriptions, Hoover Institution, Stanford University, Stanford, CA 94305-6010

Nonprofit
U.S. Postage Paid

Permit No. 26

Peterborough, NH

03458

****************AUTO**5-DIGIT 94301
10070 XXXX 109 #
Mr. Ron K Unz 266 #
No. 371
555 Bryant St
Palo Alto CA 94301-1704